

Estimation of Seemingly Unrelated Tobit Regressions via the EM Algorithm

Cliff J. Huang and Frank A. Sloan

Department of Economics, Vanderbilt University, Nashville, TN 37235

Killard W. Adamache

Department of Health Economics, Virginia Commonwealth University, Richmond, VA 23298

An expectation-maximum (EM) likelihood algorithm is used to estimate two seemingly unrelated Tobit regressions in which the dependent variables are truncated normal. An illustrative example on the determination of the life-health insurance and pension benefits is also given.

KEY WORDS: Truncated normal; Conditional expectation; Maximum likelihood estimation; Sufficient statistics; Latent variables; Censored models.

1. INTRODUCTION

In this article we consider the estimation of two seemingly unrelated Tobit regressions in which the dependent variables are truncated normal. The model is useful, since it can be viewed as the reduced form of a simultaneous-equations Tobit model. The proposed estimation method and algorithm are interesting in themselves for the following reasons. In the estimation of a simultaneous equations model, for example, Nelson and Olson (1978) proposed a procedure analogous to the two-stage least squares method. In the first stage of estimating the reduced form, however, the disturbances are assumed uncorrelated; hence some iterative algorithms of maximum likelihood or instrumental-variable techniques apply to each of the reduced-form equations separately. A more efficient first-stage estimate of the truncated dependent variables can be obtained by taking into account the nonzero covariance between the disturbances, even though the regressors are identical in the different Tobit regressions.

In this article, the expectation-maximization (EM) algorithm of Dempster, Laird, and Rubin (1977) is applied to compute the maximum likelihood estimates in the case of nonzero covariance. We then provide an illustrative example on the determination of life-health insurance and pension benefits.

2. THE MODEL AND THE EM ALGORITHM

Consider the estimation of two seemingly unrelated Tobit regressions,

$$\begin{bmatrix} Z_{1t} \\ Z_{2t} \end{bmatrix} = \begin{bmatrix} X_{1t} & 0 \\ 0 & X_{2t} \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} + \begin{bmatrix} U_{1t} \\ U_{2t} \end{bmatrix}$$

or

$$Z_t = X_t B + U_t, \quad t = 1, 2, \dots, n, \quad (1)$$

where Z_t is a vector of random latent variables and the

matrix of exogenous variables X_t is assumed to be non-stochastic. The observed data Y_{it} are related to the latent variable Z_{it} by the observation function $Y(Z_{it})$,

$$\begin{aligned} Y_{it} &= Y(Z_{it}) = Z_{it} \quad \text{iff } Z_{it} > 0 \\ &= 0 \quad \text{iff } Z_{it} \leq 0. \end{aligned} \quad (2)$$

The vector of disturbances, $U'_t = (U_{1t} \ U_{2t})$, is assumed to be distributed according to $N(0, \Sigma)$, where

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

and $E(U_{is} \ U_{js'}) = 0$ for $s \neq s'$.

Naturally, the model (1), along with the observation function (2), can be considered as the reduced form of a simultaneous-equations model, such as the model of Amemiya (1974) or the model of Nelson and Olson (1978). The model of seemingly unrelated Tobit regression (SUTR) considered here conforms to the reduced form of the latter model, but not to Amemiya's. The reduced form of the Amemiya model is different in the sense that the observation function (2) is replaced by $Y_{it} = Y(Z_{it}) = Z_{it}$ if all observed dependent variables are positive and $Y_{it} = 0$ otherwise.

Let $\theta = (B, \Sigma)$ be the unknown parameter set. The likelihood function of the latent variables $Z = (Z_1, \dots, Z_n)$ is

$$\begin{aligned} L(Z; \theta) &= (2\pi)^{-n} |\Sigma|^{-n/2} \\ &\times \exp \left\{ -\frac{1}{2} \left[\sum_i \sum_j \sigma^{ij} \left(\sum_{t=1}^n Z_{it} Z_{jt} \right) \right. \right. \\ &- 2 \sum_i \sum_j \sigma^{ij} \left(\sum_{t=1}^n X_{it} Z_{jt} \right) B_i \\ &\left. \left. + \sum_i \sum_j \sigma^{ij} B'_i \left(\sum_{t=1}^n X'_{it} X_{jt} \right) B_j \right] \right\}, \end{aligned} \quad (3)$$

where σ^{ij} is the element of inverse matrix Σ^{-1} . Since $(\sum_{i=1}^n Z_{it}Z_{jt})$ and $(\sum_{i=1}^n X_{it}Z_{jt})$ are joint sufficient statistics for θ , the likelihood function $L(Z; \theta)$ can be more compactly written as

$$L(Z; \theta) = K(Z)\exp[P'(\theta)t(Z)] \cdot [A(\theta)]^{-1}, \quad (4)$$

where $P(\theta)$ is a vector function of θ and $t(Z)$ is the corresponding vector of sufficient statistics.

Corresponding to $L(Z; \theta)$, the marginal likelihood function of the observed data $Y = (Y_1, \dots, Y_n)$ is

$$L(Y; \theta) = \int_R L(Z; \theta) dZ, \quad (5)$$

where $R = \{Z; Y(Z) = Y\}$ is the set of latent variables given the observed data Y . Maximization of the log-likelihood of (5),

$$\begin{aligned} \log L(Y; \theta) &= \log \int_R K(Z)\exp[P'(\theta)t(Z)] dZ - \log A(\theta), \end{aligned}$$

results in the following first-order condition:

$$\begin{aligned} d \log A(\theta)/d\theta &= [dP'(\theta)/d\theta] \left\{ \int_R \exp[P'(\theta)t(Z)] dZ \right\}^{-1} \\ &\quad \times \int_R t(Z)\exp[P'(\theta)t(Z)] dZ \\ &= [dP'(\theta)/d\theta] \cdot E[t(Z) | Y; \theta], \end{aligned} \quad (6)$$

where $E[t(Z) | Y; \theta]$ is the conditional expectation of the sufficient statistics.

On the other hand, if the latent variables Z were completely observed, the maximization of the log-likelihood function in (4),

$\log L(Z; \theta) = \log K(Z) + P'(\theta)t(Z) - \log A(\theta)$, would simply yield the condition that

$$\frac{d \log A(\theta)}{d\theta} = \frac{dP'(\theta)}{d\theta} t(Z). \quad (7)$$

Thus, by comparing the maximization conditions (6) and (7), Dempster et al. (1977) suggested an iterative procedure for obtaining the maximum likelihood estimate of θ . Suppose $\hat{\theta}_{(p)}$ denotes the current estimated values of θ after p iterations of the algorithm. The next iteration involves two steps. In the expectation step (the E step), the conditional expectation of the sufficient statistics on the right side of (6) is evaluated at $\hat{\theta}_{(p)}$; that is, $E[t(Z) | Y; \hat{\theta}_{(p)}]$. The estimated sufficient statistics are then used to replace $t(Z)$ in the log-likelihood function $\log L(Z; \theta)$ for the maximization purpose (the M step). The maximization condition (7) is used for solving the next iterative value $\hat{\theta}_{(p+1)}$; that is,

$$\frac{d \log A(\hat{\theta}_{(p+1)})}{d\theta} = \frac{dP'(\hat{\theta}_{(p)})}{d\theta} E[t(Z) | Y; \hat{\theta}_{(p)}]. \quad (8)$$

Given the new estimate $\hat{\theta}_{(p+1)}$, the iteration continues between the expectation and the maximization steps until the convergence occurs.

3. CONDITIONAL EXPECTATION OF THE SUFFICIENT STATISTICS

Let the set of integers $\{1, 2, \dots, n\}$ be divided into four mutually exclusive and exhaustive subsets, S_i ($i = 1, 2, 3, 4$), such that

$$S_1 = \{t | Y_{1t} > 0, Y_{2t} > 0\}$$

$$S_2 = \{t | Y_{1t} = 0, Y_{2t} > 0\}$$

$$S_3 = \{t | Y_{1t} > 0, Y_{2t} = 0\}$$

$$S_4 = \{t | Y_{1t} = 0, Y_{2t} = 0\}.$$

The conditional expectation of the sufficient statistics $(\sum_{i=1}^n Z_{it}Z_{jt})$ and $(\sum_{i=1}^n X_{it}Z_{jt})$, given the observed data Y , can then be decomposed into four groups accordingly. They are

$$E \left[\sum_{i=1}^n Z_{it}Z_{jt} | Y; \theta \right] = \sum_{g=1}^4 \sum_{t \in S_g} E[Z_{it}Z_{jt} | Y; \theta] \quad (9)$$

and

$$E \left[\sum_{i=1}^n X_{it}Z_{jt} | Y; \theta \right] = \sum_{g=1}^4 \sum_{t \in S_g} X_{it}E[Z_{jt} | Y; \theta]. \quad (10)$$

Obviously, for the observations that belong to the subset S_1 , the conditional expectations are simply

$$\sum_{t \in S_1} E[Z_{it}Z_{jt} | Y; \theta] = \sum_{t \in S_1} Y_{it}Y_{jt}$$

and

$$\sum_{t \in S_1} X_{it}E[Z_{jt} | Y; \theta] = \sum_{t \in S_1} X_{it}Y_{jt}.$$

For the observations belonging to other subsets, S_2 , S_3 , and S_4 , the conditional expectations need to be evaluated. Define the standardized random variable of Z_{it} ,

$$T_{it} = (Z_{it} - X_{it}B_i)/\sigma_i,$$

and the corresponding standardized truncated point at $Z_{it} = 0$,

$$h_{it} = -X_{it}B_i/\sigma_i.$$

For observations belonging to the subset S_2 , the latent variable Z_{1t} is truncated from above at 0, $Z_{1t} \leq 0$, and Z_{2t} is not truncated. Correspondingly, the standardized variable T_{1t} is truncated from above as h_{1t} (T_{2t} is not truncated). The conditional moments of T_{1t} in this case are defined as

$$\begin{aligned} M(r) &= E \left[T_{1t}^r | T_{1t} < h_{1t}, T_{2t} = \frac{Y_{2t} - X_{2t}B_2}{\sigma_2} \right], \\ & \quad r = 1, 2, 3, \dots \quad (11) \end{aligned}$$

The first two moments are shown to be (Johnson and Kotz 1970, p. 113)

$$M(1) = \rho T_{2t} - \frac{\phi(\alpha_{2t})}{\Phi(\alpha_{2t})} \sqrt{1 - \rho^2}$$

$$M(2) = (1 - \rho^2) \left\{ 1 - \frac{\alpha_{2t} \phi(\alpha_{2t})}{\Phi(\alpha_{2t})} - \left[\frac{\phi(\alpha_{2t})}{\Phi(\alpha_{2t})} \right]^2 \right\} + [M(1)]^2, \quad (12)$$

where

$$\rho = \sigma_{12} / \sqrt{\sigma_{11} \sigma_{22}}, \quad \alpha_{1t} = (h_{2t} - \rho T_{1t}) / \sqrt{1 - \rho^2},$$

$$\alpha_{2t} = (h_{1t} - \rho T_{2t}) / \sqrt{1 - \rho^2},$$

and $\phi(\cdot)$ and $\Phi(\cdot)$ are standard normal density and distribution functions, respectively. The conditional expectation for $t \in S_2$ in (9) and (10) are then

$$\sum_{t \in S_2} E[Z_{1t} Z_{2t} | Y_t; \theta] = \sum_{t \in S_2} E[Z_{1t} | Y_t; \theta] Y_{2t}$$

$$= \sum_{t \in S_2} [X_{1t} B_1 + \sigma_1 M(1)] Y_{2t}$$

$$\sum_{t \in S_2} E[Z_{1t}^2 | Y_t; \theta]$$

$$= \sum_{t \in S_2} [(X_{1t} B_1)^2 + 2 X_{1t} B_1 \sigma_1 M(1) + \sigma_1^2 M(2)]$$

$$\sum_{t \in S_2} X_{it} E[Z_{1t} | Y_t; \theta] = \sum_{t \in S_2} X_{it} [X_{1t} B_1 + \sigma_1 M(1)]. \quad (13)$$

For the observations belonging to the subset S_3 , the conditional expectations are similarly obtained from (11)–(13) by interchanging the subscripts 1 and 2.

Lastly, for the observations belonging to the subset S_4 , both Z_{1t} and Z_{2t} are truncated from above at 0. The moments of the standardized variables T_{1t} and T_{2t} are defined as

$$M(r, s) = E[T_{1t}^r T_{2t}^s | T_{1t} \leq h_{1t}, T_{2t} \leq h_{2t}]. \quad (14)$$

The first two moments are

$$M(1, 0) = -\frac{1}{\Phi(h_{1t}, h_{2t}; \rho)}$$

$$\times [\phi(h_{1t}) \Phi(A_{1t}) + \rho \phi(h_{2t}) \Phi(A_{2t})]$$

$$M(2, 0) = \frac{1}{\Phi(h_{1t}, h_{2t}; \rho)} [-h_{1t} \phi(h_{1t}) \Phi(h_{1t})$$

$$- \rho^2 h_{2t} \phi(h_{2t}) \Phi(h_{2t})$$

$$+ \rho(1 - \rho^2) \phi(h_{1t}, h_{2t}; \rho)] + 1$$

$$M(1, 1) = \frac{1}{\Phi(h_{1t}, h_{2t}; \rho)} \{ \rho [-h_{1t} \phi(h_{1t}) \Phi(A_{1t})$$

$$- h_{2t} \phi(h_{2t}) \Phi(A_{2t}) + \Phi(h_{1t}, h_{2t}; \rho)]$$

$$+ (1 - \rho^2) \phi(h_{1t}, h_{2t}; \rho) \},$$

where

$$A_{1t} = (h_{2t} - \rho h_{1t}) / \sqrt{1 - \rho^2},$$

$$A_{2t} = (h_{1t} - \rho h_{2t}) / \sqrt{1 - \rho^2},$$

and $\phi(h_{1t}, h_{2t}; \rho)$ and $\Phi(h_{1t}, h_{2t}; \rho)$ are bivariate standard normal density and distribution functions. The moments for $M(0, 1)$ and $M(0, 2)$ are obtained from those $M(1, 0)$ and $M(2, 0)$ by interchanging h_{1t} and h_{2t} , and A_{1t} and A_{2t} . The conditional expectations for $t \in S_4$ are then

$$\sum_{t \in S_4} E[Z_{1t} | Y_t; \theta] = \sum_{t \in S_4} [X_{1t} B_1 + \sigma_1 M(1, 0)],$$

$$\sum_{t \in S_4} E[Z_{1t}^2 | Y_t; \theta]$$

$$= \sum_{t \in S_4} [(X_{1t} B_1)^2 + 2 X_{1t} B_1 \sigma_1 M(1, 0) + \sigma_1^2 M(2, 0)],$$

and

$$\sum_{t \in S_4} E[Z_{1t} Z_{2t} | Y_t; \theta]$$

$$= \sum_{t \in S_4} [(X_{1t} B_1)(X_{2t} B_2) + X_{1t} B_1 \sigma_2 M(0, 1)$$

$$+ X_{2t} B_2 \sigma_1 M(1, 0) + \sigma_1 \sigma_2 M(1, 1)].$$

The conditional expectations $E[Z_{2t} | Y_t; \theta]$ and $E[Z_{2t}^2 | Y_t; \theta]$ are obtained by interchanging the subscripts 1 and 2.

4. ESTIMATION VIA THE EM ALGORITHM

Suppose that $\hat{\theta}_{(p)}$ denotes the current estimated value of θ after p iterations of the algorithm. In the E step, the sufficient statistics $(\sum_{t=1}^n Z_{it} Z_{jt})$ and $(\sum_{t=1}^n X_{it} Z_{jt})$ in (3) are estimated by $(\sum_{t=1}^n Z_{it} Z_{jt})_{(p)}$ and $(\sum_{t=1}^n X_{it} Z_{jt})_{(p)}$ from (9) and (10):

$$\left(\sum_{t=1}^n Z_{it} Z_{jt} \right)_{(p)} = \sum_{g=1}^4 \sum_{t \in S_g} E[Z_{it} Z_{jt} | Y; \hat{\theta}_{(p)}]$$

and

$$\left(\sum_{t=1}^n X_{it} Z_{jt} \right)_{(p)} = \sum_{g=1}^4 \sum_{t \in S_g} X_{it} E[Z_{jt} | Y; \hat{\theta}_{(p)}].$$

These estimated sufficient statistics are then used for the maximization step of the algorithm. In the M step, the new estimate $\hat{\theta}_{(p+1)}$ of θ is obtained by maximizing the log-likelihood function (3) to obtain

$$\hat{B}_{(p+1)} = \begin{bmatrix} \hat{\sigma}_{(p)}^{11} \sum_{t=1}^n X'_{1t} X_{1t} & \hat{\sigma}_{(p)}^{12} \sum_{t=1}^n X'_{1t} X_{2t} \\ \hat{\sigma}_{(p)}^{21} \sum_{t=1}^n X'_{2t} X_{1t} & \hat{\sigma}_{(p)}^{22} \sum_{t=1}^n X'_{2t} X_{2t} \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} \hat{\sigma}_{(p)}^{11} \left(\sum_{t=1}^n X'_{1t} Z_{1t} \right)_{(p)} + \hat{\sigma}_{(p)}^{12} \left(\sum_{t=1}^n X'_{1t} Z_{2t} \right)_{(p)} \\ \hat{\sigma}_{(p)}^{21} \left(\sum_{t=1}^n X'_{2t} Z_{1t} \right)_{(p)} + \hat{\sigma}_{(p)}^{22} \left(\sum_{t=1}^n X'_{2t} Z_{2t} \right)_{(p)} \end{bmatrix}, \quad (15)$$

where $[\hat{\sigma}_{ij}^p]$ is the inverse matrix of Σ evaluated at the p th iteration. The new estimate of $(\hat{\sigma}_{ij})_{(p+1)}$ is then

$$\begin{aligned}
 &(\hat{\sigma}_{ij})_{(p+1)} \\
 &= \frac{1}{n} \left[\left(\sum_{t=1}^n Z_{it} Z_{jt} \right)_{(p)} - \hat{B}'_i \left(\sum_{t=1}^n X'_{it} Z_{jt} \right)_{(p)} \right. \\
 &\quad \left. - \hat{B}'_j \left(\sum_{t=1}^n X'_{jt} Z_{it} \right) + \hat{B}'_i \left(\sum_{t=1}^n X'_{it} X_{jt} \right) \hat{B}_j \right]. \quad (16)
 \end{aligned}$$

Any consistent estimator of θ can be used as a starting value for the iteration. The illustrative example given in Section 5 uses the consistent estimator of Amemiya (1974, p. 1010) based on the instrumental variable technique.

Louis (1982) proposed a procedure of obtaining the observed information matrix when using the EM algorithm. It requires a computation of the gradient and second derivatives matrix of the logarithmic likelihood function $\ln L(Z; \theta)$ under the restriction $Z \in R = \{Z; Y(Z) = Y\}$. From (5), the gradient vector of $\ln L(Y; \theta)$ is

$$\begin{aligned}
 \frac{d \ln L(Y; \theta)}{d\theta} &= \int_R \frac{dL(Z; \theta)}{d\theta} dZ \Big/ \int_R L(Z; \theta) dZ \\
 &= \left\{ E \left[\frac{d \ln L(Z; \theta)}{d\theta} \right] \Big| Z \in R \right\}, \quad (17)
 \end{aligned}$$

where the last equality is obtained by multiplying and dividing the integrand in the numerator by $L(Z; \theta)$.

Taking an additional derivative, the second derivative matrix can be shown to be

$$\begin{aligned}
 \frac{d^2 \ln L(Y; \theta)}{d\theta d\theta'} &= E \left[\frac{d^2 \ln L(Z; \theta)}{d\theta d\theta'} \Big| Z \in R \right] \\
 &+ E \left[\left(\frac{d \ln L(Z; \theta)}{d\theta} \right) \left(\frac{d \ln L(Z; \theta)}{d\theta} \right)' \Big| Z \in R \right] \\
 &- \left\{ E \left[\frac{d \ln L(Z; \theta)}{d\theta} \Big| Z \in R \right] \right\} \\
 &\times \left\{ E \left[\frac{d \ln L(Z; \theta)}{d\theta} \Big| Z \in R \right] \right\}'.
 \end{aligned}$$

When the first-order condition of the maximization is imposed, Equation (17) is 0. Thus, at the last iteration of the EM algorithm, the last term of the second derivative is 0. The second derivative matrix of the observed data Y used in the calculation of the covariance matrix of $\hat{\theta} = (\hat{B}, \hat{\Sigma})$ can then be computed in the last iteration of the EM algorithm using only the first two terms.

Clearly the EM approach is simply an alternative algorithm for the maximum likelihood estimation of θ in (5) via the “complete-data” specification $L(Z; \theta)$. Various other algorithms such as the Newton method and the Davidon–Fletcher–Powell method are com-

monly used to find the optimal value of θ directly via the “incomplete-data” specification $L(Y; \theta)$. Unfortunately, tests of the performance of these iterative algorithms are seldom investigated and compared in the literature. This certainly is an interesting and important research subject to be explored in the future.

5. AN ILLUSTRATIVE EXAMPLE

The share of compensation paid in nonwage form has risen dramatically in the last three decades. The increase has been substantial for both legally mandated and nonmandated fringes. Excluding various types of leaves, the largest nonmandated fringes by far are health insurance and private pensions. In a recent empirical study, Sloan and Adamache (1986) assessed variations in fringes among establishments that paid in the form of life-health insurance and pension per work hour. The analysis was based largely on a standard economic theory in modeling household behavior in which the employee values both nonwage and wage income. The employees’ choice between the two depends, among other factors, on federal and state marginal tax rates and family income. Other explanatory variables include union status of the establishment; a dummy variable indicating whether the establishment’s product was subject to regulation (such as in the electric utility, gas, communication, and transportation industries); a dummy variable for manufacturing industries; and the capital-to-labor ratio for the establishment’s industry, because turnover tends to be lower in firms with a high ratio, which makes the demand for pensions higher. Several explanatory variables also are included to account for differences in employee characteristics—proportions of workers who are female, under age 44, over age 55, and part-time, as well as variables for metropolitan status, region, and year.

The principal data sources were Surveys of Employer Expenditure for Employee Compensation (EEC) conducted by the U.S. Bureau of Labor Statistics in 1968, 1972, and 1977. The dependent variables were employer contributions to life-health insurance and private pension plans expressed per work hour. The EEC data base did not allow life insurance to be split from health insurance, but more than eight-tenths of the composite is for health insurance. Out of a sample of 30,686 observations on establishments, more than 51% (15,737) of them provided no pension and more than 16% (5,063) provided neither life-health nor pension benefit.

Based on the estimation of two separate Tobit regressions for life-health insurance and private pensions, Sloan and Adamache (1986) concluded that both nonwage compensations are highly responsive to the marginal income tax rate. Tax elasticities at the mean for both nonwage fringes exceed 1. No account, however, was taken of likely positive correlation of the disturbance terms in the two Tobit regressions. We have, therefore, reestimated the two seemingly unrelated To-

bit regressions based on the EM algorithm of the maximum likelihood method.

Table 1 shows the estimates of SUTR via the EM algorithm and the estimates of unrelated Tobit regression (UTR) reported by Sloan and Adamache (1986). The correlation coefficient (ρ) between the disturbances in the two Tobit regressions is estimated to be .492, which is significantly larger than 0 based on an asymptotic test ($t = 123$). The nonzero correlation coefficient implies the possible gain of efficiency in the SUTR estimation. As shown in both life-health and pension regressions, the SUTR estimates all have smaller estimated asymptotic variances than the UTR estimates, although both the SUTR and UTR estimations produce valid asymptotic standard errors, since the regressions place no restrictions on coefficients. The reduction in asymptotic variance in the life-health regression is

smaller than in the pension regression. This is expected, however, because the fraction of limited observations (i.e., 0 values) in the life-health insurance dependent variable is much smaller than the fraction of limited observations in the pension dependent variable. The fractions are 16.5% in the former and 51.3% in the latter.

Technically, the comparison between the SUTR and UTR estimations is valid only with respect to the possible gain of efficiency caused by the existence of nonzero correlation between disturbances in the Tobit regressions. It is the asymptotic variances of the estimators that are of most interest. Nevertheless, noticing the substantial change in the estimates of the marginal tax rate in the pension regression is inescapable. The SUTR estimate shows the effect of the marginal tax rate to be about half as large as the UTR estimate. The

Table 1. Nonwage Compensation per Hour Regressions: Comparison of Two Approaches

Explanatory variable	Life-health regressions		Pension regressions	
	SUTR approach	UTR approach	SUTR approach	UTR approach
Marginal tax rate	2.270 ^a (.199)	2.35 ^a (.20)	1.116 ^a (.428)	2.28 ^a (.56)
Family income (in thousands of dollars)	-.0011 (.0013)	-.0020 (.0013)	.046 ^a (.003)	.041 ^a (.004)
Unionized establishment	.168 ^a (.003)	.17 ^a (.003)	.295 ^a (.006)	.35 ^a (.009)
Regulated industry	.017 ^a (.005)	.017 ^a (.005)	-.012 (.011)	-.094 ^a (.015)
Manufacturing industry	.060 ^a (.003)	.060 ^a (.003)	.017 ^a (.007)	.022 ^b (.009)
Capital-labor ratio	.0002 (.0005)	.0001 (.0006)	.018 ^a (.001)	.017 ^a (.001)
Office worker	.064 ^a (.005)	.074 ^a (.005)	.015 (.011)	.10 ^a (.016)
Proportion female workers	-.153 ^a (.006)	-.16 ^a (.007)	.071 ^a (.013)	.0053 (.020)
Proportion minority workers	-.021 (.019)	-.014 (.019)	.305 ^a (.041)	.54 ^a (.06)
Proportion workers under age 24	-.099 ^a (.018)	-.096 ^a (.018)	-.084 ^a (.038)	-.16 ^a (.06)
Proportion workers over age 55	.068 ^a (.021)	.091 ^a (.021)	-.171 ^a (.046)	-.055 (.066)
Proportion workers part-time	-.182 ^a (.011)	-.20 ^a (.011)	-.314 ^a (.025)	-.38 ^a (.035)
Standard metropolitan statistical area	.036 ^a (.003)	.038 ^a (.003)	.087 ^a (.006)	.11 ^a (.008)
Northeast	.046 ^a (.003)	.043 ^a (.003)	.092 ^a (.007)	.10 ^a (.009)
North Central	.063 ^a (.003)	.063 ^a (.003)	.080 ^a (.006)	.11 ^a (.009)
West	.070 ^a (.004)	.071 ^a (.004)	.082 ^a (.008)	.11 ^a (.011)
1972 dummy	.059 ^a (.004)	.079 ^a (.004)	.805 ^a (.008)	.97 ^a (.014)
1977 dummy	.096 ^a (.007)	.097 ^a (.007)	.923 ^a (.016)	.94 ^a (.024)
Constant	-.302 ^a (.022)	-.32 ^a (.02)	-1.845 (.047)	-2.15 ^a (.065)
Variance σ_{ϵ}	.038 ^a (.0003)		.174 ^a (.001)	
Correlation coefficient ρ	.492 ^a (.004)		.492 ^a (.004)	

NOTE: Asymptotic standard errors are in parentheses.

^a Statistically significant at the 1% level.

^b Statistically significant at the 5% level.

SUTR estimate of tax elasticity evaluated at the means is well under unity, but the associated income elasticity is only slightly higher than the corresponding UTR estimate. The estimated union effect is somewhat more modest in the SUTR version, and the estimated impact of several of the demographic variables is changed—up in some cases and down in others. In this case (but not for life-health insurance in which there are relatively few values of the dependent variable at the limit), it certainly paid to take account of the correlated disturbances.

6. CONCLUSION

In this article we have considered the estimation of two seemingly unrelated Tobit regressions via the EM algorithm. We have argued that a more efficient estimation is as obtainable in the censored SUTR model as in the case of Zellner's (1962) uncensored SUR model. Several problems have been omitted from the discussion, however, and will be investigated in future research. First, the proposed SUTR estimation through the EM algorithm is restricted to a model with only two equations. The obvious reason is a practical as well as a theoretical one. In a model with more than two equations, the conditional expectations of the sufficient statistics are difficult, if not impossible, to obtain. Second, in this article the censored SUTR and SUR estimators are compared only in terms of the estimated variances. The efficiency comparison ignores the predictive performance of the estimators. Perhaps the cross-validation technique with the estimation obtained from the

first subsample by the SUTR and SUR methods should be used to compare the predictive performance of the censored dependent variables for the second subsample. This split-sample analysis deserves careful consideration as a future research topic in the area of a censored regression model.

ACKNOWLEDGMENT

We would like to thank a referee and the editors for their helpful comments and corrections.

[Received October 1985. Revised November 1986.]

REFERENCES

- Amemiya, Takeshi (1974), "Multivariate Regression and Simultaneous Equation Models When the Dependent Variables Are Truncated Normal," *Econometrica*, 42, 999–1012.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Ser. B*, 39, 1–22.
- Johnson, Norman L., and Kotz, Samuel (1970), *Distributions in Statistics: Continuous Multivariate Distribution*, New York: John Wiley.
- Louis, Thomas A. (1982), "Finding the Observed Information Matrix When Using the EM Algorithm," *Journal of the Royal Statistical Society, Ser. B*, 44, 226–233.
- Nelson, Forrest, and Olson, Lawrence (1978), "Specification and Estimation of a Simultaneous-Equation Model With Limited Dependent Variables," *International Economic Review*, 19, 695–704.
- Sloan, Frank A., and Adamache, Killard W. (1986), "Taxation and the Growth of Nonwage Compensation," *Public Finance Quarterly*, 14, 115–137.
- Zellner, A. (1962), "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association*, 57, 348–368.