Semiparametric Estimation and Inference in a System of Censored Demand Equation

by

Rafic Fahs, Scott Cardell, and Ron Mittelhammer[1]

2001 AAEA Annual Conference
Chicago, IL

Abstract:

The purpose of this paper is to utilize the generalized method of moments (GMM) approach for estimating a system of multivariate Tobit equations and propose a practical consistent estimator of model parameters. The GMM approach is based on a common set of general marginal and bivariate moment relations that hold between explanatory variables and model noise.

---

[1] Rafic Fahs is a Ph.D. candidate in Agricultural Economics at Washington State University. Scott Cardell and Ron Mittelhammer are an Adjunct Professor and Professor, respectively in the Department of Agricultural Economics and also in the Program in Statistics, Washington State University.

1. Introduction

In this paper we consider a simultaneous equation model where the dependent variables follow a truncated multivariate normal distribution and develop an alternative econometric methodology for estimating a system of censored demand equations based on the GMM approach. The model in this paper is a generalization of univariate Tobit models to a system of Tobit-type equations exhibiting contemporaneous noise correlation.

The multivariate Tobit-type model has wide applicability in economic analyses, including such applications as:

1. consumer demand derived from utility maximization over convex budget sets subject to non-negativity constraints,
2. input demands under non-negativity constraints or quantity rationing,
3. production under non-negativity constraint or quotas, and
4. labor supply of consumers.

We refer the reader to Lee (1993) for a brief survey of multivariate Tobit models.

In a systems approach, the censored regression equations typically have correlated error terms and require a joint estimation procedure, which generally leads to the specification of a mixed distribution consisting of a continuous probability density function for the positive observations and a discrete probability mass function for the zero observations Lee (1993). In this case the estimation of the multivariate model by classical maximum likelihood methods requires the evaluation of partially integrated multivariate normal probability density function, which is known to be computationally inefficient, progressively more inaccurate, or even intractable as the dimensionality of the integral problem increases much beyond three. The maximum likelihood method involves the evaluation of these integrals at each iteration of a maximization algorithm, where the dimensionality of the integral is equal to the dimension of the discontinuous part.

To avoid the computational problems affecting the maximum likelihood method, Pudney (1989) estimated a system of Tobit equations one equation at a time (marginally), by applying the Tobit technique to each equation in turn. While numerically tractable and consistent, such estimators tend to be inefficient because they totally ignore the intercorrelation between the equations and do not impose any cross equation restrictions that may apply. Amemiya (1974) considered a model that is based only on the positive sample outcomes of *all* of the dependent variables jointly. The merit of his estimation procedure include consistency and numerical tractability, as in Pudney's method, and also takes into account intercorrelation of the error components of the model. But his model is not applicable to all situations, and in particular would be very inefficient if the numbers of sample observations that resulted in all positive-valued dependent variable values is very small, which would be the case if the probability of joint positivity is small.

Maddala (1976b) modified Amemiya's (1973) procedure so as to use all the sample observations pertaining to the model, but a major shortcoming of his procedure is the evaluation of partially integrated multivariate normal probability density functions with all the attendant problems indicated previously. Given the problem of high dimensional integration when evaluating the probability of the discrete (often zero) observations, researchers have historically often restricted their attention to special classes of multivariate Tobit models that were computationally tractable or that ignored the censored observations. Ignoring the specific data sampling characteristics of the zero observations when modeling and estimating the parameters of the model will lead to biased estimates. Excluding the censored observations will result in sample selection bias (Lee and Pitt (1987)). To cope with the preceding issues, the GMM approach can be applied by specifying a common set of general marginal and bivariate moment relations that hold between explanatory variables and model noise under the tenets of the censored regression model.

For our specific empirical problem in which the moment information is stated and given, the GMM procedure makes asymptotically optimal use of the information. Regarding the number of estimating equations or moments to use, asymptotic efficiency comparisons suggest more is better than less. The asymptotic covariance matrix of the GMM estimator generally becomes smaller (by a positive definite matrix) as the number of estimating equations used increases.

Bivariate moments help to identify and estimate the parameters involved in the covariance structure occurring across equation errors that would not be estimable if only marginal univariate moments were used. In addition, the bivariate moments avoid the problem of evaluating the probability of the discontinuous part in higher dimensions because numerical integration is then only required in two dimensions, which is quite accurate and fast computationally. Including both the bivariate and marginal univariate moments also adds to the efficiency of the estimation method.

The remainder of the paper is organized as follows: section 2 reviews the univariate Tobit model and the marginal moments associated with it. In section 3, we extend the univariate Tobit to a system of Tobit-type equations and derive the first two moments and the moment generating function of the truncated multivariate normal distribution. In section 4, we introduce generalized method of moments (GMM) estimator applied to the multivariate Tobit model. In section 5, the results of Monte Carlo experiments are presented and analyzed. The concluding remarks in section 6 summarize the major implications of the paper.

2. Standard Univariate Tobit (Censored Regression) Model

Consider a regression model

$$Y_i^* = \mathbf{x}_{i.}\beta + \varepsilon_i, \quad i = 1,2,\ldots.n,$$

where β is $K\times 1$ column vector of unknown parameters; $\mathbf{x}_{i.}$ Is a$1\times K$ row vector of explanatory variable values; and the ε_i's are residuals that are independently and normally distributed with mean zero and a common variance σ^2. Let $\theta = (\beta, \sigma)$ denotes the parameters of the model (2.1).

Suppose Y_i^* is not observed and instead we observe Y_i defined as

$$\begin{aligned} Y_i &= Y_i^* \quad \text{if } Y_i^* > 0 \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{2.1}$$

This model is the traditional univariate Tobit model. The likelihood function for this model is

$$L = \prod_{y_i=0}[1-F(\mathbf{x}_{i.}\beta/\sigma)]\prod_{y_i>0}[\sigma^{-1}f((y_i-\mathbf{x}_{i.}\beta)/\sigma)] \tag{2.2}$$

where F and f are the cumulative distribution and density functions of the standard normal distribution, respectively. If one observes neither y_i nor $\mathbf{x}_{i.}$ when $y_i^* \leq 0$, the model is known as a truncated regression model whose likelihood function is

$$L = \prod_{y_i>0} [F(\mathbf{x}_{i.}\beta/\sigma)]^{-1}[\sigma^{-1} f((y_i - \mathbf{x}_{i.}\beta)/\sigma)] \quad (2.3)$$

Note that (2.2) can be written as

$$L = \prod_{y_i=0} [1 - F(\mathbf{x}_{i.}\beta/\sigma)] \prod_{y_i>0} F(\mathbf{x}_{i.}\beta/\sigma) \prod_{y_i>0} [F(\mathbf{x}_{i.}\beta/\sigma)^{-1} \sigma^{-1} f((y_i - \mathbf{x}_{i.}\beta)/\sigma)] \quad (2.4)$$

The first two products on the right-hand side of (2.4) are equivalent to the likelihood function of a Probit model, and the last product term is the likelihood function of the truncated Tobit model as given in (2.3).

Considering the model given by equation (2.1) and conditioning on the positive observations, the first and second conditional moments of a truncated normal random variable are given by

$$E(Y_i \mid Y_i > 0) = \mathbf{x}_{i.}\beta + \sigma \frac{f}{F} \quad (2.5)$$

$$E(Y_i^2 \mid Y_i > 0) = \left(\mathbf{x}_{i.}\beta\right) E(Y_i \mid Y_i > 0) + \sigma^2 \quad (2.6)$$

where henceforth we are assuming that f and F are shorthand for $f(\mathbf{x}_{i.}\beta/\sigma)$ and $F(\mathbf{x}_{i.}\beta/\sigma)$, respectively. If we use all of the observations on y_i, instead of using only the nonzero observations, the unconditional first and second moments are given by

$$E(Y_i) = \left(\mathbf{x}_{i.}\beta\right) F + \sigma \cdot f \quad (2.7)$$

$$E(Y_i^2) = \left(\mathbf{x}_{i.}\beta\right) E(Y_i) + \sigma^2 \cdot F \quad (2.8)$$

Now define a binary variable, analogous to the Probit case, as

$$Y_{BINi} = \begin{cases} 1 & \text{if } y_i > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then another moment condition can be defined by noting that

$$E(Y_{BINi}) = F \, . \tag{2.9}$$

Given all of the moment conditions (2.5)-(2.9) for all n observations, we can define the following relationships between **Y**, **X**, and disturbances $\mu^{(i)}$ that have zero expectations:

$$\mathbf{Y}_{>\mathbf{0}} = \mathbf{X}_{>\mathbf{0}}\beta + \sigma \odot \frac{1}{\mathbf{F}_{>\mathbf{0}}} \odot \mathbf{f}_{>\mathbf{0}} + \mu^{(1)} \tag{2.10}$$

$$\mathbf{Y}^{\mathbf{2}}_{>\mathbf{0}} = \mathbf{X}_{>\mathbf{0}}\beta \odot E\mathbf{Y}_{>\mathbf{0}} + \sigma^2 + \mu^{(2)} \tag{2.11}$$

$$\mathbf{Y} = \mathbf{X}\beta \odot \mathbf{F} + \sigma \odot \mathbf{f} + \mu^{(3)} \tag{2.12}$$

$$\mathbf{Y}^2 = \mathbf{X}\beta \odot \mathbf{Y} + \sigma^2 \odot \mathbf{F} + \mu^{(4)} \tag{2.13}$$

and

$$\mathbf{Y}_{BIN} = \mathbf{F} + \mu^{(5)} \tag{2.14}$$

where $\mathbf{Y}_{>\mathbf{0}}$ denotes the set of positive valued observations, $\mathbf{X}_{>\mathbf{0}}$ denotes the observations on the explanatory variables that correspond to the positive valued outcomes, $E\mathbf{Y}_{>\mathbf{0}}$ is shorthand notation for $E\mathbf{Y}_{>\mathbf{0}} \equiv E\left(\mathbf{Y} \mid \mathbf{Y} > \mathbf{0}\right) = \mathbf{X}_{>\mathbf{0}}\beta + \sigma \odot \frac{1}{\mathbf{F}_{>\mathbf{0}}} \odot \mathbf{f}_{>\mathbf{0}}$ and recall that $\odot$ denotes the Hadamard (elementwise) product. In this context, **F** and **f** are *vectors* of cumulative distribution function and density function values of the N(0,1) distribution, evaluated at the vector $\mathbf{X}\beta \odot \left(\frac{1}{\sigma}\right)$, $\mathbf{F}_{>\mathbf{0}}$ and $\mathbf{f}_{>\mathbf{0}}$ are the subsets of those vectors corresponding to the positive valued observations $\mathbf{Y}_{>0}$, and $\frac{1}{\mathbf{F}_{>0}}$ denotes a vector of reciprocals of the elements in $\mathbf{F}_{>0}$.

Because

$$\mu^{(1)} = \mathbf{Y}_{>0} - \mathbf{X}_{>0}\beta - \sigma \odot \frac{1}{\mathbf{F}_{>0}} \odot \mathbf{f}_{>0},$$

$$\mu^{(2)} = \mathbf{Y}^{2}_{>0} - \mathbf{X}_{>0}\beta \odot E\mathbf{Y}_{>0} - \sigma^{2},$$

$$\mu^{(3)} = \mathbf{Y} - \mathbf{X}\beta \odot \mathbf{F} - \sigma \odot \mathbf{f},$$

$$\mu^{(4)} = \mathbf{Y}^{2} - \mathbf{X}\beta \odot \mathbf{Y} - \sigma^{2} \odot \mathbf{F},$$

and

$$\mu^{(5)} = \mathbf{Y}_{\mathrm{BIN}} - \mathbf{F},$$

orthogonality conditions can be defined as follows:

$$\mathrm{E}\left[\mathbf{X}'_{>0}(\mathbf{Y}_{>0} - \mathbf{X}_{>0}\beta - \sigma \odot \frac{1}{\mathbf{F}_{>0}} \odot \mathbf{f}_{>0})\right] = \mathbf{0},$$

$$\mathrm{E}\left[\mathbf{X}'_{>0}(\mathbf{Y}^{2}_{>0} - \mathbf{X}_{>0}\beta \odot E\mathbf{Y}_{>0} - \sigma^{2})\right] = \mathbf{0},$$

$$\mathrm{E}\left[\mathbf{X}'(\mathbf{Y} - \mathbf{X}\beta \odot \mathrm{F} - \sigma \odot \mathbf{f})\right] = \mathbf{0},$$

$$\mathrm{E}\left[\mathbf{X}'(\mathbf{Y}^{2} - \mathbf{X}\beta \odot \mathbf{Y} - \sigma^{2} \odot \mathbf{F})\right] = \mathbf{0},$$

and

$$\mathrm{E}\left[\mathbf{X}'(\mathbf{Y}_{\mathrm{BIN}} - \mathbf{F})\right] = \mathbf{0}.$$

We can then define a $(5K \times 1)$ vector of moment conditions derived from the orthogonality conditions as

$$\mathrm{E}\left[\mathbf{h}(\mathbf{Y},\mathbf{X},\theta)\right]=\mathrm{E}\begin{bmatrix} \mathbf{X}_{>0}'(\mathbf{Y}_{>0}-\mathbf{X}_{>0}\beta-\ \sigma\odot\dfrac{1}{\mathbf{F}_{>0}}\odot\mathbf{f}_{>0}) \\ \mathbf{X}_{>0}'(\mathbf{Y}_{>0}^{2}-\mathbf{X}_{>0}\beta\odot E\mathbf{Y}_{>0}-\sigma^{2}) \\ \mathbf{X}'(\mathbf{Y}-\mathbf{X}\beta\odot\mathrm{F}-\sigma\odot\mathbf{f}) \\ \mathbf{X}'(\mathbf{Y}^{2}-\mathbf{X}\beta\odot\mathbf{Y}-\sigma^{2}\odot\mathbf{F}) \\ \mathbf{X}'(\mathbf{Y}_{\mathrm{BIN}}-\mathbf{F}) \end{bmatrix}=\ \mathbf{0}. \tag{2.20}$$

The sample analog of the population moments displayed in (2.20) is

$$\mathbf{h}(\mathbf{y},\mathbf{x},\theta)=\begin{bmatrix} \dfrac{\mathbf{x}_{>0}'}{\mathrm{n}^{(1)}}(\mathbf{y}_{>0}-\mathbf{x}_{>0}\beta-\ \sigma\odot\dfrac{1}{\mathbf{F}_{>0}}\odot\mathbf{f}_{>0}) \\ \dfrac{\mathbf{x}_{>0}'}{\mathrm{n}^{(2)}}(\mathbf{y}_{>0}^{2}-\mathbf{x}_{>0}\beta\odot E\mathbf{y}_{>0}-\sigma^{2}) \\ \dfrac{\mathbf{x}'}{\mathrm{n}^{(3)}}(\mathbf{y}-\mathbf{x}\beta\odot\mathbf{F}-\sigma\odot\mathbf{f}) \\ \dfrac{\mathbf{x}'}{\mathrm{n}^{(4)}}(\mathbf{y}^{2}-\mathbf{x}\beta\odot\mathbf{y}-\sigma^{2}\odot\mathbf{F}) \\ \dfrac{\mathbf{x}'}{\mathrm{n}^{(5)}}(\mathbf{y}_{\mathrm{BIN}}-\mathbf{F}) \end{bmatrix}=\ \mathbf{0} \tag{2.21}$$

where $n^{(i)}$ denotes the number of sample observations that correspond to the ith set of moment conditions and $E\mathbf{y}_{>0}$ is equal to $E\mathbf{Y}_{>0}$ evaluated at sample outcomes for **y** and **x** and at specified values for β and σ. Note that a sufficient condition for the orthogonality conditions to be valid is that

$$\mathrm{E}\left[\mu^{(\nu)}\,|\,\mathbf{X}_{>0}\right]=\mathbf{0}\ \text{ for }\nu=1,2$$

$$\mathrm{E}\left[\mu^{(\nu)}\,|\,\mathbf{X}\ \right]=\mathbf{0}\ \text{ for }\nu=3,4,5\ . \tag{2.22}$$

The dimension of $\mathbf{h}(\mathbf{Y},\mathbf{X},\theta)$ is $5\mathrm{K}\times 1$, which is greater than the number of unknown parameters in the model. Consequently, in general there will not exist a unique parameter vector θ that solves the sample moment conditions via the ordinary method of moments approach, which attempts to find a θ that satisfies (2.21). We will deal with solution issues in section 4 ahead.

3. Multivariate Tobit Model

In a multivariate Tobit model we have

$$\mathbf{Y}_t^* = \mathbf{X}_t\beta + \varepsilon_t, \quad t = 1,...,n$$

where $\varepsilon_t \sim N(\mathbf{0}, \Sigma)$ and $\mathbf{Y}_t^* \sim N(\mathbf{X}_t\beta, \Sigma)$ and t denotes the observation number. In this model, $\mathbf{Y}_t^*$ is a $J \times 1$ continuous latent variables that determine the choice probabilities of an optimizing agent,

$$\mathbf{X}_t\beta = \begin{bmatrix} X_{1t}\beta_1 \\ X_{2t}\beta_2 \\ . \\ . \\ . \\ X_{Jt}\beta_J \end{bmatrix}$$

, where $\mathbf{X}_{jt}$ is $1 \times k_j$ row vector of explanatory variables**,** β_j is $k_j \times 1$ column vector of parameters**,** and Σ is a $J \times J$ variance covariance matrix. Assuming the censoring occurs at zero, the model is generally written as

$$Y_{jt} = \begin{cases} Y_{jt}^* & \text{if } Y_{jt}^* > 0 \\ 0 & \text{otherwise} \end{cases} \quad t = 1,2,\ldots.n, \;\; j = 1,2,\ldots\ldots.J \tag{3.1}$$

where the subscript t denotes observation and subscript j denotes choice alternatives.

In general for a given individual decision maker, the vector $\mathbf{Y}_t^*$ can be written as

$\mathbf{Y}_{\mathbf{t}}^{*} = \begin{bmatrix} \mathbf{Y}_{\mathrm{td}}^{*} \\ \mathbf{Y}_{\mathrm{tc}}^{*} \end{bmatrix}$, where $\mathbf{Y}_{\mathrm{td}}^{*}$ represents the discrete elements of $\mathbf{Y}_{\mathrm{t}}^{*}$ associated with the outcomes $\mathbf{Y}_{\mathrm{td}} = \mathbf{0}$ and $\mathbf{Y}_{\mathrm{tc}}^{*}$ represents the noncensored observations corresponding to $\mathbf{Y}_{\mathrm{tc}} = \mathbf{Y}_{\mathrm{tc}}^{*} > [\mathbf{0}]$.

Let $\theta = [\beta, \Sigma]$ denote all of the parameters of the model. Then the joint probability density of ($\mathbf{Y}_{\mathrm{td}}^{*}, \mathbf{Y}_{\mathrm{tc}}^{*}$) can be represented in general by

$$\mathrm{p}(\mathbf{Y}_{\mathrm{td}}^{*}, \mathbf{Y}_{\mathrm{tc}}^{*};\ \mathbf{X}, \theta) = \mathrm{p}(\mathbf{Y}_{\mathrm{td}}^{*} \mid \mathbf{Y}_{\mathrm{tc}}^{*};\ \mathbf{X}, \theta) \cdot \mathrm{p}(\mathbf{Y}_{\mathrm{tc}}^{*};\ \mathbf{X}, \theta),$$

where $\mathrm{p}(\mathbf{Y}_{\mathrm{td}}^{*} \mid \mathbf{Y}_{\mathrm{tc}}^{*};\ \mathbf{X}, \theta)$ represents the discrete choice probability conditional on the continuous variable $\mathbf{Y}_{\mathrm{tc}}^{*}$, and $\mathrm{p}(\mathbf{Y}_{\mathrm{tc}}^{*};\ \mathbf{X}, \theta)$ is the probability density function for the continuous variable. Thus the contribution of the t$^{\mathrm{th}}$ observation to the likelihood function will be

$$\mathrm{L}(\mathbf{Y}_{\mathrm{td}}^{*}, \mathbf{Y}_{\mathrm{tc}}^{*};\ \mathbf{X}, \theta) = \mathrm{L}(\mathbf{Y}_{\mathrm{td}}^{*} \mid \mathbf{Y}_{\mathrm{tc}}^{*};\ \mathbf{X}, \theta) \cdot \mathrm{L}(\mathbf{Y}_{\mathrm{tc}}^{*};\ \mathbf{X}, \theta) \qquad (3.2)$$

3.1 Moments of the Truncated Bivariate Normal distribution

The first two moments and the moment generating function of the truncated multivariate normal distribution have been derived by Tallis(1961). Amemiya (1974) found an interesting relationship between the first and second moments. Adapted from Tallis and Amemiya, and based on a different parameterization of the model, we derive the first and second moments of the truncated bivariate normal distribution. We present the result for the pair of random variables $(\varepsilon_1, \varepsilon_2)$, but the subscripts can be changed to i and j without consequence, *mutatis mutandis*.

Theorem : Let the density of $\varepsilon = (\varepsilon_1, \varepsilon_2)'$ be given by

$$\begin{cases} \frac{1}{P} f(\varepsilon_1,\ \varepsilon_2) & \text{for } \varepsilon_i > a_i,\ i = 1,2 \\ 0 & \text{elsewhere,} \end{cases} \tag{3.1.1}$$

where f is the density of the bivariate standard normal distribution and P is defined by

$P = \int_{a_1}^{\infty}\int_{a_2}^{\infty} f(\tau_1,\ \tau_2)\ d\tau_1\ d\tau_2$. Then

$$P \cdot E(\varepsilon_t) = \sum_{j=1}^{2} \rho_{tj}\ f(a_j) \left[1\text{-}F\left(C\left(j,\tau(j)\right)\right)\right],\quad \tau(j) \equiv \{1,2\} - j, \tag{3.1.2}$$

and

$$P \cdot E(\varepsilon_1\varepsilon_2) = \rho_{12}P + \sum_{j=1}^{2} \rho_{j1}\rho_{j2}a_j\ \ f(a_j) \left[1\text{-}F\left(C\left(j,\tau(j)\right)\right)\right]$$
$$+ \sum_{j=1}^{2}\left\{\rho_{j1} \sum_{r\neq j} \sqrt{\rho_{r2} - \rho_{jr}\rho_{j2}} \cdot f(a_j) \cdot f\left(\gamma_{jr}\right)\right\}, \tag{3.1.3}$$

where $C\left(j,\tau(j)\right) = \dfrac{a_{\tau(j)} - \rho_{\tau(j)j}a_j}{\sqrt{1\text{-}\rho^2_{\tau(j)j}}}$ and $\gamma_{jr} = \dfrac{a_r - \rho_{rj}a_j}{\sqrt{1\text{-}\rho_{rj}{}^2}}$.

Proof: See the Appendix.

It is useful for establishing moment conditions below to also note that Amemiya (1974) found an interesting relationship between the first and second moments that can be represented as

$$\rho^{i'}\ E\left[\varepsilon_i\varepsilon\right] = 1 + a_i\rho^{i'}E\left[\varepsilon\right] \tag{3.1.4}$$

where $\rho^{i'}$ is the ith row of $\mathbf{R}^{-1}$, where $\mathbf{R}$ is the correlation matrix of the error vector.

Now note in general that in a multivariate Tobit model characterized by J alternative choices, there are $\frac{J(J-1)}{2}$ alternative pairs of decision outcomes that can be examined in a bivariate manner. For example, in a three choice model, there are 3 bivariate pairs for each observation given by $(y_{1t}, y_{2t}),(y_{1t}, y_{3t}),(y_{2t}, y_{3t})$. For each pair, we can derive scalar first and second moments of one of the variables in the pair conditional on an event for the other, resulting in two moment definitions per pair, and we can also establish a cross-product moment for each pair. All of these results can be derived based on (3.1.2) and (3.1.3). Writing the moments in terms of covariances, and letting $a = (a_1, a_2)' = (\mathbf{x}_{jt}\beta_j, \mathbf{x}_{kt}\beta_k)'$ in (3.1.1), we have the following first and second moment relations:

$$E(\mathbf{y}_j^2 | \mathbf{y}_j > 0, \mathbf{y}_k > 0) = \sigma_{jk}^* + \frac{\sigma_{jk}}{\sigma_{kk}} \odot \mathbf{y}_k \odot \mathbf{y}_k + \mathbf{x}_j\beta_j \odot \mathbf{y}_j - \frac{\sigma_{jk}}{\sigma_{kk}} \odot \mathbf{x}_k\beta_k \odot \mathbf{y}_k \qquad (3.1.5)$$

$$E(\mathbf{y}_j | \mathbf{y}_j > 0, \mathbf{y}_k > 0) = \mathbf{x}_j\beta_j + \frac{1}{P} \odot \sum_{l \in \{j,k\}} \frac{\sigma_{jl}}{\sigma_{ll}} \odot f(a_l) \odot \left[1 - F(C_{li}^*)\right] \qquad (3.1.6)$$

$$\begin{aligned} E(\mathbf{y}_j\mathbf{y}_k | \mathbf{y}_j > \mathbf{0}, \mathbf{y}_k > \mathbf{0}) = {} & \mathbf{x}_k\beta_k \odot E(\mathbf{y}_j | \mathbf{y}_j > \mathbf{0}) + \mathbf{x}_j\beta_j \odot E(\mathbf{y}_k | \mathbf{y}_k > \mathbf{0}) - \mathbf{x}_j\beta_j \odot \mathbf{x}_k\beta_k \\ & + \sigma_{jk} + \frac{1}{P} \odot \sum_{l \in \{j,k\}} \frac{\sigma_{lj}\sigma_{lk}}{\sigma_{ll}^2} \cdot a_l^* \odot f(a_l^*) \odot [1 - F(C_{lj}^*)] \qquad (3.1.7) \\ & + \frac{1}{P} \odot \sum_{l \in \{j,k\}} \left\{ \frac{\sigma_{lj}\sqrt{\sigma_{kk}}}{\sigma_{ll}} \odot \sum_{r \neq l} \left[\frac{\sigma_{lr} - \frac{\sigma_{lr}\sigma_{lk}}{\sigma_{ll}}}{\sqrt{\sigma_{ll}\sigma_{rr} - \sigma_{lr}^2}} \right] \odot f(a_l^*) \odot f(\gamma_{lr}^*) \right\}, \end{aligned}$$

where $a_1^* = \mathbf{x}_j\beta_j \odot \frac{1}{\sqrt{\sigma_{jj}}}$, $a_2^* = \mathbf{x}_k\beta_k \odot \frac{1}{\sqrt{\sigma_{kk}}}$, $\sigma_{li}^* = \sqrt{\sigma_{ii} - \frac{\sigma_{il}^2}{\sigma_{ll}}}$, $\sigma_{lr}^* = \sqrt{\sigma_{rr} - \frac{\sigma_{lr}^2}{\sigma_{ll}}}$,

$C_{li}^* = \dfrac{a_i - \frac{\sigma_{li}}{\sigma_{ll}} a_l}{\sigma_{li}^*}$ with $l \in \{j,k\}$ and $i = \{j,k\} - l$, and $\gamma_{lr} = \dfrac{a_r - \frac{\sigma_{lr}}{\sigma_{ll}} a_l}{\sigma_{lr}^*}$ $(r \neq l)$.

We can rewrite (3.1.5), (3.1.6), and (3.1.7) in terms of equations relating **Y** with **X** and a zero-mean residual term $\mu^{(i)}$ as, respectively,

$$\mathbf{y}_j^2 = \sigma_{jk}^* + \frac{\sigma_{jk}}{\sigma_{kk}} \odot \mathbf{y}_k \odot \mathbf{y}_k + \mathbf{x}_j\beta_j \odot \mathbf{y}_j - \frac{\sigma_{jk}}{\sigma_{kk}} \odot \mathbf{x}_k\beta_k \odot \mathbf{y}_k + \mu^{(6)} \tag{3.1.8}$$

$$\mathbf{y}_j = \mathbf{x}_j\beta_j + \frac{1}{P} \odot \sum_{l \in \{j,k\}} \frac{\sigma_{jl}}{\sigma_{ll}} \odot f(a_l) \odot \left[1 - F(C_{li}^*)\right] + \mu^{(7)} \tag{3.1.9}$$

and letting

$$\begin{aligned}\mathbf{z} = {} & \mathbf{x}_k\beta_k \odot \mathbf{y}_j + \mathbf{x}_j\beta_j \odot \mathbf{y}_k - \mathbf{x}_j\beta_j \odot \mathbf{x}_k\beta_k \\ & + \sigma_{jk} + \frac{1}{P} \odot \sum_{l \in \{j,k\}} \frac{\sigma_{lj}\sigma_{lk}}{\sigma_{ll}^2} \cdot a_l^* \odot f(a_l^*) \odot [1 - F(C_{li}^*)] \\ & + \frac{1}{P} \odot \sum_{l \in \{j,k\}} \left\{ \frac{\sigma_{lj}\sqrt{\sigma_{kk}}}{\sigma_{ll}} \odot \sum_{r \neq l} \left[\frac{\sigma_{lr} - \frac{\sigma_{lr}\sigma_{lk}}{\sigma_{ll}}}{\sqrt{\sigma_{ll}\sigma_{rr} - \sigma_{lr}^2}} \right] \odot f(a_l^*) \odot f(\gamma_{lr}^*) \right\}\end{aligned} \tag{3.1.10}$$

then

$$\mathbf{y}_j\mathbf{y}_k = \mathbf{z} + \mu^{(8)}. \tag{3.1.11}$$

3.2 Estimating Equations Based on Bivariate Moments

Similar to the orthogonality conditions presented in section 2 relating to the marginal moments, and noting that for any pair of decision outcomes, (y_j, y_k), there are two moments of the type (3.1.8) and (3.1.9) and one moment of type (3.1.11), we can define a

$\frac{5J(J-1)K}{2} \times 1$ vector function of bivariate-type moment conditions based on the preceding results as

$$E\left[\mathbf{hBIV}(\mathbf{Y},\mathbf{X},\theta)\right] = \begin{bmatrix} \mathbf{X}'\left(\mathbf{Y}_j^{2} - \sigma_{jk}^{*} - \frac{\sigma_{jk}}{\sigma_{kk}} \odot \mathbf{Y}_k \odot \mathbf{Y}_k - \mathbf{X}_j\beta_j \odot \mathbf{Y}_j + \frac{\sigma_{jk}}{\sigma_{kk}} \odot \mathbf{X}_k\beta_k \odot \mathbf{Y}_k\right) \\ \cdot \\ \cdot \\ \mathbf{X}'\left(\mathbf{Y}_j - \mathbf{X}_j\beta_j - \frac{1}{P} \odot \sum_{l\in\{j,k\}} \frac{\sigma_{jl}}{\sigma_{ll}} \odot f(a_l) \odot \left[1\text{-}F(C_{li}^{*})\right]\right) \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{X}'(\mathbf{Y}_j\mathbf{Y}_k - \mathbf{Z}) \end{bmatrix} = [\mathbf{0}] \tag{3.2.1}$$

The sample analog of the population moment condition (3.2.1) is

$$\mathbf{hBIV}(\mathbf{y},\mathbf{x},\theta) = \begin{bmatrix} \frac{\mathbf{x}'}{n^{(6)}}\left(\mathbf{y}_j^{2} - \sigma_{jk}^{*} - \frac{\sigma_{jk}}{\sigma_{kk}} \odot \mathbf{y}_k \odot \mathbf{y}_k - \mathbf{x}_j\beta_j \odot \mathbf{y}_j + \frac{\sigma_{jk}}{\sigma_{kk}} \odot \mathbf{x}_k\beta_k \odot \mathbf{y}_k\right) \\ \cdot \\ \cdot \\ \frac{\mathbf{x}'}{n^{(7)}}\left(\mathbf{y}_j - \mathbf{x}_j\beta_j - \frac{1}{P} \odot \sum_{l\in\{j,k\}} \frac{\sigma_{jl}}{\sigma_{ll}} \odot f(a_l) \odot \left[1\text{-}F(C_{li}^{*})\right]\right) \\ \cdot \\ \cdot \\ \cdot \\ \frac{\mathbf{x}'}{n^{(8)}}(\mathbf{y}_j\mathbf{y}_k - \mathbf{z}) \end{bmatrix} = [\mathbf{0}] \tag{3.2.2}$$

where as before, $n^{(i)}$ denotes the number of sample observations involved in the ith set of moment conditions. The number of equations in the vector $\mathbf{hBIV}(\mathbf{Y}, \mathbf{X}, \theta)$ is $\frac{5J(J-1)K}{2}$, which is greater than the number of unknown parameters in the model, $\frac{J(J-1)}{2} + K$.

Consequently, there will generally not exist a unique parameter vector θ that solves the sample moment conditions via the ordinary method of moments approach, which attempts to find a θ that satisfies (3.2.2). We will deal with such issues in section 4 ahead.

3.3. Comparison To Moment Conditions of Amemiya and Maddala

In general, sample observations corresponding to any multivariate Tobit model consisting of J distinct choices can be indicate any of 2^J alternative choice combinations. For example, in a two choice model.

$$Y_{1i} = \begin{cases} \mathbf{x}_{1i}\beta_1 + \varepsilon_{1i} & \text{if RHS} > 0 \\ 0 & \text{if RHS } \leq 0 \end{cases}$$

$$Y_{2i} = \begin{cases} \mathbf{x}_{2i}\beta_2 + \varepsilon_{2i} & \text{if RHS} > 0 \\ 0 & \text{if RHS } \leq 0 \end{cases}$$

and we can classify an observation to be one of $2^2 = 4$ possibilities. Divide the sample observations into 4 sets:

$$S_1 = \{i: y_{1i} > 0,\ y_{2i} > 0\}$$

$$S_2 = \{i: y_{1i} > 0,\ y_{2i} = 0\}$$

$$S_3 = \{i: y_{1i} = 0,\ y_{2i} > 0\}$$

$$S_4 = \{i: y_{1i} = 0,\ y_{2i} = 0\}$$

Letting f represent the density of the bivariate normal random vector (ε_{1i}, ε_{2i}), we have the following likelihood function for

$$L = \prod_{i \in S_1} f\left(Y_{1i} - \mathbf{x}_{1i}\beta_1,\ Y_{2i} - \mathbf{x}_{2i}\beta_2\right) \prod_{i \in S_2} \int_{-\infty}^{-\mathbf{x}_{2i}\beta_2} f\left(Y_{1i} - \mathbf{x}_{1i}\beta_1,\ \varepsilon_{2i}\right) d\varepsilon_{2i}$$
$$\prod_{i \in S_3} \int_{-\infty}^{-\mathbf{x}_{1i}\beta_1} f\left(\varepsilon_{1i}, Y_{2i} - \mathbf{x}_{2i}\beta_2\right) d\varepsilon_{1i} \prod_{i \in S_4} \int_{-\infty}^{-\mathbf{x}_{1i}\beta_1} \int_{-\infty}^{-\mathbf{x}_{2i}\beta_2} f\left(\varepsilon_{1i},\ \varepsilon_{2i}\right) d\varepsilon_{1i} d\varepsilon_{2i}$$

where $\prod_{i \in S_j}$ denotes the product over all observations in the set S_j.

Amemiya (1974) uses only observations in S_1 and suggests an indirect least square procedure based on evaluating the first and second moments of y_{1i} and y_{2i} in S_1. His model for a two choice model begins with the following specification:

$$\begin{cases} Y_{jt} = \mathbf{x}_{jt}\beta_j + \varepsilon_{jt} \\ Y_{kt} = \mathbf{x}_{kt}\beta_j + \varepsilon_{kt} \end{cases} \quad \text{if } Y_{jt} > 0 \text{ and } Y_{kt} > 0$$
$$Y_{jt} = Y_{kt} = 0 \qquad \text{otherwise.}$$

Amemiya (1974) uses the moment derived previously for the truncated bivariate normal and obtains

$$\begin{aligned} E(Y_{1i} \mid S_1) &= \mathbf{x}_{1i}\beta_1 + \frac{1}{P}\sigma_1 f_1(1-F_2) + \frac{1}{P}\frac{\sigma_{12}}{\sigma_2} f_2(1-F_1) \\ E(Y_{2i} \mid S_1) &= \mathbf{x}_{2i}\beta_2 + \frac{1}{P}\sigma_2 f_2(1-F_1) + \frac{1}{P}\frac{\sigma_{12}}{\sigma_1} f_1(1-F_2) \end{aligned} \tag{3.3.1}$$

where

f_1 is the density of N(0,1) evaluated at $\frac{\mathbf{x}_{1i}\beta_1}{\sigma_1}$,

f_2 is the density of N(0,1) evaluated at $\frac{\mathbf{x}_{2i}\beta_2}{\sigma_2}$,

F_1 is the cumulative distribution of N(0,1) evaluated at $\frac{-\mathbf{x}_{1i}\beta_1 + \frac{\sigma_{12}}{\sigma_2^2}\mathbf{x}_{2i}\beta_2}{\sigma_1^*}$,

F_2 is the cumulative distribution of N(0,1) evaluated at $\frac{-\mathbf{x}_{2i}\beta_2 + \frac{\sigma_{12}}{\sigma_1^2}\mathbf{x}_{1i}\beta_1}{\sigma_2^*}$,

$P = \text{Prob}(\varepsilon_{1i} > \frac{-\mathbf{x}_{1i}\beta_1}{\sigma_1}, \varepsilon_{2i} > \frac{-\mathbf{x}_{2i}\beta_2}{\sigma_2})$,

$\sigma_1^{*2} = \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2}$ and $\sigma_2^{*2} = \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}$.

The equations in (3.3.1) can be rewritten as

$$Y_{1i} = \mathbf{x}_{1i}\beta_1 + \frac{1}{P}\sigma_1 f_1(1-F_2) + \frac{1}{P}\frac{\sigma_{12}}{\sigma_2} f_2(1-F_1) + \mu_{1i}$$
$$Y_{2i} = \mathbf{x}_{2i}\beta_2 + \frac{1}{P}\sigma_2 f_2(1-F_1) + \frac{1}{P}\frac{\sigma_{12}}{\sigma_1} f_1(1-F_2) + \mu_{2i} \qquad (3.3.2)$$

where $E(u_{1i}) = E(u_{2i}) = 0$.

The equation (3.3.2) can be estimated by ordinary least squares provided estimates of P, f_1, f_2, F_1, and F_2 are available. Amemiya (1974) suggests obtaining an initial consistent estimates in the first step by using an instrumental method applied to the following moments

$$Y_{1i}^{\ 2} = \sigma_1^{*2} + \frac{\sigma_{12}}{\sigma_2^{\ 2}} Y_{1i}Y_{2i} + \mathbf{x}_{1i}\beta_1 Y_{1i} - \frac{\sigma_{12}}{\sigma_2^{\ 2}}\mathbf{x}_{2i}\beta_2 Y_{1i} + \omega_{1i}$$
$$Y_{2i}^{\ 2} = \sigma_2^{*2} + \frac{\sigma_{12}}{\sigma_1^{\ 2}} Y_{1i}Y_{2i} + \mathbf{x}_{2i}\beta_2 Y_{2i} - \frac{\sigma_{12}}{\sigma_1^{\ 2}}\mathbf{x}_{1i}\beta_1 Y_{1i} + \omega_{2i} \qquad (3.3.3)$$

and then obtains estimates of (3.3.2) by OLS in a second step. There are two main difficulties regarding the implementation of the procedure that Amemiya suggests. First of all, there may be cases in which the subset S_1 consists of very few sample observations compared with the total number of sample observations available. Secondly the estimates of P become slow, inaccurate, and eventually intractable as the dimensionality of the choice problem (the size of J) , and concurrently the dimensionality of the censored observations increase.

Maddala (1976b) modifies Amemiya's (1974) procedure so as to use all of the sample observations. The first step in his procedure is analogous to Amemiya (1974). In his second step he defines P_1, P_2, and P_3 to be the probabilities that the observations belong to sets S_1, S_2, and S_3, respectively. Then equation (3.3.2) can be rewritten as

$$\begin{aligned} E\left[Y_{1i}\right] &= P_1(\mathbf{x}_{1i}\beta_1) + P_2(\mathbf{x}_{1i}\beta_1) + P_1\ E\left[Y_{1i} \mid S_1\right] + P_2\ E\left[Y_{1i} \mid S_2\right] \\ E\left[Y_{2i}\right] &= P_1(\mathbf{x}_{2i}\beta_2) + P_3(\mathbf{x}_{2i}\beta_2) + P_1\ E\left[Y_{2i} \mid S_2\right] + P_3\ E\left[Y_{2i} \mid S_3\right] \end{aligned} \qquad (3.3.4)$$

The short coming of his procedure is that an estimate of the P_i's is still required, which becomes slow, inaccurate, and eventually intractable as the dimensionality of the choice problem and concurrently the dimensionality of the censored observations increase.

It follows that the development of alternative methods for estimating a system of censored demand equations is a highly relevant endeavor.

4. GMM Estimation of the Tobit System

In this section a GMM approach is applied to the Tobit system by utilizing the general marginal and bivariate moment relations that were derived in sections 2 and 3.1-

3.2. We can define a $5KJ\times 1$ vector of marginal moment conditions, based on results in section 2, as

$$\mathrm{E}\left[\mathbf{hMAR}(\mathbf{Y},\mathbf{X},\theta)\right]=\mathrm{E}\begin{bmatrix}\mathbf{X}'_{>0}(\mathbf{Y}_{>0}-\mathbf{X}_{>0}\beta-\ \sigma\odot\frac{1}{\mathbf{F}_{>0}}\odot\mathbf{f}_{>0}) \\ \mathbf{X}'_{>0}(\mathbf{Y}^{2}_{>0}-\mathbf{X}_{>0}\beta\odot E\mathbf{Y}_{>0}-\sigma^{2}) \\ \mathbf{X}'(\mathbf{Y}-\mathbf{X}\beta\odot\mathrm{F}-\sigma\odot\mathbf{f}) \\ \mathbf{X}'(\mathbf{Y}^{2}-\mathbf{X}\beta\odot\mathbf{Y}-\sigma^{2}\odot\mathbf{F}) \\ \mathbf{X}'(\mathbf{Y}_{\mathrm{BIN}}-\mathbf{F})\end{bmatrix}=\mathbf{0}. \tag{4.1}$$

The sample analog of the population moment condition (4.1) is

$$\mathbf{hMAR}(\mathbf{y},\mathbf{x},\theta)=\begin{bmatrix}\frac{\mathbf{x}'_{>0}}{\mathrm{n}^{(1)}}(\mathbf{y}_{>0}-\mathbf{x}_{>0}\beta-\ \sigma\odot\frac{1}{\mathbf{F}_{>0}}\odot\mathbf{f}_{>0}) \\ \frac{\mathbf{x}'_{>0}}{\mathrm{n}^{(2)}}(\mathbf{y}^{2}_{>0}-\mathbf{x}_{>0}\beta\odot E\mathbf{y}_{>0}-\sigma^{2}) \\ \frac{\mathbf{x}'}{\mathrm{n}^{(3)}}(\mathbf{y}-\mathbf{x}\beta\odot\mathbf{F}-\sigma\odot\mathbf{f}) \\ \frac{\mathbf{x}'}{\mathrm{n}^{(4)}}(\mathbf{y}^{2}-\mathbf{x}\beta\odot\mathbf{y}-\sigma^{2}\odot\mathbf{F}) \\ \frac{\mathbf{x}'}{\mathrm{n}^{(5)}}(\mathbf{y}_{\mathrm{BIN}}-\mathbf{F})\end{bmatrix}=\mathbf{0} \tag{4.2}$$

Regarding the bivariate moments, the sample analog of the bivariate population moment condition (3.2.1) is given by $\mathbf{hBIV}(\mathbf{Y},\mathbf{X},\theta)$, as defined in (3.2.2) of section 3. Then a cumulative set of moment conditions, including both marginal and bivariate moments, can be specified as

$$\mathrm{E}\mathbf{H}(\mathbf{Y},\mathbf{X},\theta)=\mathrm{E}\begin{bmatrix}\mathbf{hMAR}(\mathbf{Y},\mathbf{X},\theta) \\ \mathbf{hBIV}(\mathbf{Y},\mathbf{X},\theta)\end{bmatrix}=\mathbf{0}. \tag{4.3}$$

The sample estimating equation-moment analog is $\mathbf{H}(\mathbf{y},\mathbf{x},\theta)=\left[0\right]$, which has dimension $\left(5\mathrm{KJ}+\frac{5\mathrm{J}(\mathrm{J}\text{-}1)\mathrm{K}}{2}\right)\times 1$, while the parameter vector θ = (β,Σ) contains $\frac{\mathrm{J}(\mathrm{J}\text{-}1)}{2}+\mathrm{K}$ unique elements, where β is (K×1) and Σ is (J×J).

The number of unbiased estimating functions in the preceding specification is greater than the number of unknown parameters θ, so the set of estimating equation is overdetermined for estimating θ. Given $\mathbf{H}(\mathbf{y},\mathbf{x},\theta)=[0]$, under the GMM approach the parameter vector is chosen for which the sample moment conditions are as close to the zero vector as possible. The following weighted Euclidean distance is used as a measure of closeness:

$$\min_{\theta}\left[Q(\mathbf{y},\mathbf{x},\theta)\right]=\min_{\theta}\left[\ \mathbf{H}(\mathbf{y},\mathbf{x},\theta)'\,\mathbf{W}\,\mathbf{H}(\mathbf{y},\mathbf{x},\theta)\right] \qquad (4.4)$$

where $\mathbf{W}$ is a conformable positive definite symmetric weight matrix. Another way to view the GMM approach to the problem of solving overdetermined sets of sample moment conditions is through the necessary conditions for the minimization of (4.4) given by

$$\frac{\partial Q(\mathbf{y},\mathbf{x},\theta)}{\partial\theta}=2\left[\frac{\partial\mathbf{H}(\mathbf{y},\mathbf{x},\theta)}{\partial\theta}\right]'\mathbf{W}\,\mathbf{H}(\mathbf{y},\mathbf{x},\theta)=0\,. \qquad (4.5)$$

The conditions (4.5) indicate that the problem of the equation system being overdetermined is overcome by forming a $\frac{J(J\text{-}1)}{2}+K$ linear combination of the moment conditions based on the matrix $\left[\left(\frac{\partial\mathbf{H}}{\partial\theta}\right)'\right]\mathbf{W}$ to project the moment conditions to a $\frac{J(J\text{-}1)}{2}+K$-dimensional space, in effect resulting in the same number of equations as unknowns. For alternative weight matrices $\mathbf{W}$, the projection and generally the solution to the problem is altered.

The major difficulty in implementing the GMM estimator is the choice of the weighting matrix **W**, which can effect the relative efficiency of the estimator.

5. Optimum GMM weight Matrix

The choice of **W** that results in the asymptotically most efficient estimator within the class of GMM estimators, based on a given set of estimating equation-moment conditions, is the inverse of the covariance matrix $\mathbf{cov}[\mathbf{H}(\theta,\mathbf{Y},\mathbf{X})]=(\mathrm{E}(\mathbf{HH'}))^{-1}=\mathbf{W}^{*}$ (see Hansen,1982; Andrews, 1999). Optimality in the current context refers to choosing a **W** matrix in the definition of the GMM estimator

$$\hat{\theta}_{\mathrm{GMM}}(\mathbf{W}) = \arg\min_{\theta}\ [\mathbf{H}(\mathbf{Y},\mathbf{X},\theta)'\,\mathbf{W}\,\mathbf{H}(\mathbf{Y},\mathbf{X},\theta)] \qquad (5.1)$$

such that $\hat{\theta}_{\mathrm{GMM}}(\mathbf{W})$ has the smallest asymptotic covariance matrix. Because the optimal weight matrix implied by $(\mathrm{E}(\mathbf{HH'}))^{-1}=\mathbf{W}^{*}$ is generally unknown, and thus $\hat{\theta}_{\mathrm{GMM}}(\mathbf{W})$ is not operational, a consistent estimator, $\hat{\mathbf{W}}_{\mathrm{n}}$, of $\mathbf{W}^{*}$ is used. In practice, this is obtained by setting **W** = **I** (any arbitrarily chosen positive definite symmetric matrix can be chosen for **W** to obtain a consistent estimator of θ) and calculating $\hat{\theta}(\mathbf{I})$ in (5.1) and in the process calculating $\hat{\mathbf{H}}\hat{\mathbf{H}}'$. In the second step, the sample estimator of the optimal weighting matrix $\hat{\mathbf{W}}_{\mathrm{n}}$ is substituted into (5.1) leading to the estimated optimal GMM defined by $\hat{\theta}_{\mathrm{GMM}}(\hat{\mathbf{W}}_{\mathbf{n}})$. The estimated optimal GMM estimator will be consistent and asymptotically normal.

We can summarize the above procedure as follows: (1) use marginal first and second order moments for each equation, (2) use all possible combinations of bivariate first order, second order, and cross moments, (3) define the sample estimating equation analogs to the population moments, and (4) use the GMM method to estimate the model parameters.

6. Monte Carlo Results

A Monte Carlo experiment was performed on a two choice Tobit model to compare the parameter estimates of Amemiya's (1974) method of using only the observations in S_1 with the GMM procedure that includes both the marginal and bivariate moments. We used the following bivariate Tobit model:

$$\begin{bmatrix} \mathbf{y}_1^* \\ \mathbf{y}_2^* \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1\beta_1 \\ \mathbf{x}_2\beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

$$\text{where } \beta_1 = \begin{bmatrix} \beta_{11} \\ \beta_{12} \end{bmatrix} = \begin{bmatrix} .1 \\ .2 \end{bmatrix}, \beta_2 = \begin{bmatrix} \beta_{21} \\ \beta_{22} \end{bmatrix} = \begin{bmatrix} .3 \\ .4 \end{bmatrix} \text{ and } \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & .707 \\ .707 & 2 \end{bmatrix} \right). \quad (6.1)$$

The results for the model that includes both the marginal and bivariate moments as well as the results for Amemiya's method are displayed in table1. Note that the C_{ij} entries in the table refer to a reparameterization of the disturbance covariance matrix Σ into a Cholesky factorization representation, whereby **C** is a 2×2 lower triangular matrix for which $\boldsymbol{\Sigma} = \mathbf{CC}'$. In this application, the GMM weight matrix was set equal to the identity matrix. The table summarizes the results of 200 Monte Carlo repetitions and samples sizes equal to n =100, 250, and 1000 .

As expected the parameter estimates and the mean square error of the estimator that includes both the marginal and the bivariate moments outperformed the estimator based on Amemiya's method. As the sample size increases the parameters converge to the true values in table1.

Table 1. Monte Carlo Results for the 2-Choice Tobit Model[1]

Parameter	True value	GMM with marginal and bivariate moments			Amemiya's Method		
		n=100	n=250	n=1000	n=100	n=250	n=1000
β_{11}	.1	0.132	0.136	0.138	0.244	0.311	0.269
β_{12}	.2	0.187	0.193	0.193	0.192	0.190	0.200
β_{21}	.3	0.325	0.332	0.321	0.611	0.672	0.601
β_{22}	.4	0.384	0.392	0.390	0.391	0.384	0.395
C11	1	0.959	0.973	0.975	0.912	0.933	0.967
C21	.707	0.763	0.758	0.749	0.149	0.173	0.228
C22	1.224	1.150	1.164	1.183	1.225	1.241	1.284
MSE(θ)		**0.100**	**0.052**	**0.023**	**2.119**	**1.431**	**0.760**

1) The empirical means of the 200 estimator repetitions are displayed below the the sample size indicators.

We also performed a Monte Carlo experiment to estimate a trivariate Tobit model using the GMM approach that incorporates both the marginal and bivariate moments. The Tobit model was defined as

$$\begin{bmatrix} \mathbf{y}_1^* \\ \mathbf{y}_2^* \\ \mathbf{y}_3^* \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1\beta_1 \\ \mathbf{x}_2\beta_2 \\ \mathbf{x}_3\beta_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$$

$$\text{where } \beta_1 = \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \end{bmatrix} = \begin{bmatrix} 0.1 \\ 0.2 \\ 0.4 \end{bmatrix},\ \beta_2 = \begin{bmatrix} \beta_{21} \\ \beta_{22} \\ \beta_{23} \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.6 \\ 0.7 \end{bmatrix},\ \beta_3 = \begin{bmatrix} \beta_{31} \\ \beta_{32} \\ \beta_{33} \end{bmatrix} = \begin{bmatrix} 0.8 \\ 0.9 \\ 1 \end{bmatrix},$$

$$\text{and } \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} .7 & .079 & .118 \\ .079 & .9 & .201 \\ .118 & .201 & .5 \end{bmatrix} \right)$$

The covariance matrix was chosen in a way that reflects realistic variability in the error and reasonable covariance between the latent variables. Table 2 summarizes he results of 200 Monte Carlo experiments for sample sizes of n = 50, 500, 1000, and 1000. Note again that a reparameterization of the disturbance covariance matrix Σ into a Cholesky factorization representation was incorporated in estimation, where $\mathbf{C}$ is now a 3×3 lower triangular matrix for which $\mathbf{\Sigma = CC'}$.

Table 2. Monte Carlo Results for the 3-Choice Tobit Model[1]

Parameter	True value	Mean Values of the Estimates			
		N=250	N=500	N=1000	N=2000
β_{11}	.1	-0.08791989	-0.02640486	0.054318271	0.060295402
β_{12}	.2	0.21311024	0.20881330	0.20238506	0.20215332
β_{13}	.4	0.42909484	0.41805675	0.40590678	0.40306300
β_{21}	.5	0.39285850	0.47235959	0.47921165	0.48527497
β_{22}	.6	0.61435225	0.60267237	0.60450244	0.60212277
β_{23}	.7	0.71592544	0.70286628	0.70257194	0.70178116
β_{31}	.8	0.75590922	0.77931405	0.79179380	0.79479534
β_{32}	.9	0.90355379	0.90135812	0.90048774	0.90158431
β_{33}	1	1.0085011	1.0034269	1.0010977	0.99966962
C11	.83666003	0.86735759	0.87154755	0.85354373	0.84656386
C21	.09486833	0.098138846	0.099150440	0.093958745	0.080478269
C22	.94392279	0.95518833	0.93378715	0.95072268	0.95328262
C31	.14142136	0.12765288	0.13191005	0.13474203	0.14145362
C32	.19898734	0.19700645	0.14442974	0.19922134	0.19803564
C33	.66362945	0.58750382	0.64288894	0.6478198	0.65452622
MSE(θ)		**0.84015622**	**0.30500766**	**0.14269010**	**0.064219416**

1) The empirical means of the 200 estimator repetitions are displayed below the the sample size indicators.

The results for 200 Monte Carlo experiments suggest that the GMM method, based on the accumulation of marginal and bivariate moments discussed in sections 2 and 3, provides an accurate estimate of model parameters. As the sample size increases the mean square error decreases, indicative of consistency. But even for smaller sample sizes, the estimates are for the most part reasonably accurate. We also note that a some of the discrepancy in the parameters estimates may be due to the fact that we have used numerical gradients instead of analytical gradients when solving the optimization problem.

7. Conclusion

This paper has presented a methodology for estimation of a system of multivariate Tobit-type equations where the dependent variables are truncated normal. The parameter estimates of the Monte Carlo results appear quite accurate, and illustrate the potential of the proposed GMM estimation method. Estimates obtained by this procedure are consistent and asymptotically normal. Our consistent estimator may not be fully efficient, but it is an empirically tractable way of estimating a system of censored regressions involving large data sets with a relatively high dimensionality of censored observations.

Another advantage of the GMM approach is that it is not difficult to impose side constraints on the parameters that add information to the data with the potential of further increasing the precision of the estimates. For example, cross equation symmetry restrictions implied by neoclassical demand theory could be imposed by adding elements to the empirical moments vector. Furthermore, in practice there is often insufficient information to specify the parametric form of the likelihood function underlying the data sampling process. Given this situation, the GMM method in this paper can be applied to non-normally distributed data sampling processes by defining the analogs of the moment conditions used here. Overall, the principal contribution of this paper is to introduce a computationally tractable and asymptotically efficient method (when the optimal weight matrix **W** is used) of estimating a system of censored demand equations into econometric practice.

Research is currently ongoing regarding the use of analytical gradients, which will likely help to improve the accuracy of the parameter estimates and the time needed for convergence of the GMM optimization problem. We are also in the process of applying the econometric methodology to Chinese consumption data where zero outcomes are more likely to be the expression of a corner solution (rather than infrequency of purchase).

References

AMEMIYA, T. (1974) “Multivariate regression and simultaneous equation models when the dependent variable is truncated normal,” *Econometrica*, 42,999-1013.

AMEMIYA, T. (1979a) “The estimation of simultaneous equation Tobit model,” ,” *International economic review*, 20,169-181.

AMEMIYA, T. (1979b) “Tobit models: a survey. Journal of Econometrics, 24,3-61.

ANDREWS, D.W.K. (1991). “Consistent Moment Selection Procedures for Generalized Method of Moments Estimation”, *Econometrica*, 67, 543-564.

HANSEN, L. and SINGLETON (1982). “Generalized Instrumental Variable Estimators of Nonlinear Rational Expectations Models”, *Econometrica*, 50, 1269-1286.

LEE, L. (1993) “Multivariate Tobit models in econometrics” Handbook of statistics, vol. 2, 145-173.

LEE, L. AND PITT,M. (1986) “Microeconometric demand systems with binding non-negativity constrains: the dual approach, ” *Econometrica*, 54, 1237-1242.

LEE, L. AND PITT, M. (1987) “Microeconometric models of rationing, imperfect market and non-negativity constrains, ” *Journal of Econometrics*, 36, 89-110.

Maddala, G.S. (1977). “Identification and estimation problems in limited Dependent Variable Models.” In A.S.Blinder and P.Friedman (eds.), Natural Resources, Uncertainty and General Equilibirum Systems: Essays in Memory of Rafael Lusky,pp.219-39. New York: Academic.

Pudney, S. Modelling Individual Choice. Oxford: UK Basil Blackwell 1989.

TALLIS, G.M. (1961). “The Moment Generating Function of the Truncated Multi-Normal Distribution”, *Journal of the Royal Statistical Society* (Series B), 23, 223-229.

WALES, T. AND WOODLAND, A. (1983) “Consumer demand systems with binding non-negativity constrains,” Journal of econometrics, 21, 263-285.

Appendix

Theorem Proof.

Let w = (w_1, w_2) have a standard bivariate normal distribution with correlation matrix **R** and let w_i be truncated at a_i so that

$$P = \text{prob}(w_1 > a_1, w_2 > a_2) = \int_{a_1}^{\infty}\int_{a_2}^{\infty} f(\tau_1,\ \tau_2)\ d\tau_1\ d\tau_2 = F^*(a_1, a_2; \mathbf{R}).$$

The joint moment generating function corresponding to the truncated support $w_1 > a_1$, $w_2 > a_2$ is

M(t_1, t_2) =

$$M = P^{-1} \int_{a_1}^{\infty}\int_{a_2}^{\infty} e^{t'w} f(w_1,\ w_2;\ \mathbf{R})\ dw_1\ dw_2 = P^{-1}(2\pi)^{-1}\mathbf{R}^{-\frac{1}{2}}\int_{a_1}^{\infty}\int_{a_2}^{\infty} \exp[\{\mathbf{w}'\mathbf{R}^{-1}\mathbf{w}\text{-}2\mathbf{T}'\mathbf{w}\}]\ dw_1\ dw_2\,.$$

Letting **S** = **RT** where $\mathbf{T} = \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}$, then M can be rewritten as

$$M = P^{-1}(2\pi)^{-1}\mathbf{R}^{-\frac{1}{2}} e^{D}\int_{a_1}^{\infty}\int_{a_2}^{\infty} \exp\left[-\frac{1}{2}(\mathbf{w}\text{-}\mathbf{S})'\mathbf{R}^{-1}(\mathbf{w}\text{-}\mathbf{S})\right] dw_1\ dw_2,$$ where $\mathbf{D} = \frac{1}{2}\mathbf{T}'\mathbf{R}\mathbf{T}$ and

$$S_i = \sum_{j=1}^{2} \rho_{ij}\ t_j\,.$$

By the change of variables $\varepsilon_i = w_i - S_i$, and letting $b_i = a_i - S_i$, we get

$$P \cdot M = P^{-1}(2\pi)^{-1}\mathbf{R}^{-\frac{1}{2}} e^{D}\int_{\mathbf{b}_1}^{\infty}\int_{b_2}^{\infty} \exp\text{-}\frac{1}{2}\{\varepsilon'\mathbf{R}^{-1}\varepsilon\}\ d\varepsilon_1\ d\varepsilon_2 = e^{D}F^*(b_1, b_2;\ \mathbf{R})\,. \qquad \text{(A.1)}$$

Note that $f(\varepsilon_i,\ \varepsilon_j = b_j;\ \mathbf{R}) = f(b_j)\cdot f(\varepsilon_i \mid \varepsilon_j = b_j;\ \mathbf{R}_j) = f(b_i)\cdot f(z_i;\ \mathbf{R}_j)$ for $i \neq j$, where $f(b_i)$ is the marginal standard normal density of b_i, $f(z_i; \mathbf{R}_i)$ is the conditional standard normal of the ε_i variable given that the variable ε_j equals b_j,

$$R_j = 1 - \rho_{ij}^2 \text{ and } z_i = \frac{b_i - \rho_{ij} \cdot b_j}{\sqrt{1 - \rho_{ij}^2}} \; (i \neq j).$$

Note also the following

$$\frac{\partial e^D}{\partial t_i} = (t_i + \rho_{ij} t_j), \; \frac{\partial b_i}{\partial t_i} = -1, \; \frac{\partial b_i}{\partial t_j} = -\rho_{ij},$$

$$\frac{\partial F^*(b_i, b_j;\ \mathbf{R})}{\partial t_i} = \left[\frac{\partial F^*(b_i, b_j;\ \mathbf{R})}{\partial b_i}\frac{\partial b_i}{\partial t_i} + \frac{\partial F^*(b_i, b_j;\ \mathbf{R})}{\partial b_j}\frac{\partial b_j}{\partial t_i}\right],$$

$$\frac{\partial F^*(b_i, b_j;\ \mathbf{R})}{\partial b_i} = -\int_{b_j}^{\infty} f(b_i,\ \varepsilon_j)\, d\varepsilon_j = -f(b_i) \cdot \left[1 - F(z_j, R_i)\right],$$

$$\frac{\partial F^*(b_i, b_j;\ \mathbf{R})}{\partial b_j} = -\int_{b_j}^{\infty} f(\varepsilon_i, b_j)\, d\varepsilon_i = -f(b_j) \cdot \left[1 - F(z_i, R_j)\right],$$

$$\frac{\partial F^*(b_i, b_j;\ R)}{\partial t_i} = f(b_i) \cdot \left[1 - F(z_j, R_i)\right] + \rho_{ij} \cdot f(b_j) \cdot \left[1 - F(z_i, R_j)\right],$$

$$\frac{\partial z_i}{\partial t_i} = -\left(\frac{1 - \rho_{ij}^2}{\sqrt{1 - \rho_{ij}^2}}\right), \; \frac{\partial z_i}{\partial t_j} = 0, \; \frac{\partial f(b_i)}{\partial b_i} = -b_i \cdot f(b_i).$$

$$\frac{\partial^2 F^*(b_i, b_j;\ R)}{\partial t_i^2} = \frac{\partial f(b_i)}{\partial b_i} \cdot \frac{\partial b_i}{\partial t_i}\left[1\text{-}F(z_j, R_i)\ \right] +$$

$$\rho_{ij} \cdot \left[\frac{f(b_j)}{\partial b_j} \cdot \frac{\partial b_j}{\partial t_i} \cdot \left[1 - F(z_i, R_j)\right] + f(b_j) \cdot \frac{\partial\left[1\text{-}F(z_i, R_j)\right]}{\partial z_i} \cdot \frac{\partial z_i}{\partial t_i}\right].$$

Given the preceding results we can differentiate equation (A.1) first with respect to t_i and then with respect to t_j to get the following

$$P \cdot \frac{\partial M}{\partial t_i} = \frac{\partial e^D}{\partial t_i} \cdot F^*(b_i, b_j; \mathbf{R}) + \frac{\partial F^*(b_i, b_j; \mathbf{R})}{\partial t_i} \cdot e^D \text{ and}$$

$$P \cdot E(\varepsilon_i) = P \cdot \left.\frac{\partial M}{\partial t_i}\right|_{t_i = \mathbf{t_j} = 0} = f(a_i) \cdot \left[1 - F(C\left(j, \tau(j)\right); R_i)\right] + \rho_{ij} f(a_j) \cdot \left[1 - F(C\left(j, \tau(j)\right); R_j)\right], \text{ and in general}$$

$$P \cdot E(\varepsilon_t) = \sum_{j=1}^{2} \rho_{tj} \; f(a_j) \left[1\text{-}F(C\left(j, \tau(j)\right); R_j)\right]$$

and

$$P \cdot E(\varepsilon_t \varepsilon_s) = \rho_{ts} \cdot P + \sum_{j=1}^{2} \rho_{jt} \cdot \rho_{js} \cdot a_j \;\; f(a_j) \left[1\text{-}F(C\left(j, \tau(j)\right); R_j)\right]$$
$$+ \sum_{j=1}^{2} \left\{ \rho_{jt} \cdot \sum_{r \neq j} \sqrt{\rho_{rs} - \rho_{jr} \cdot \rho_{js}} \cdot f(a_j) \cdot f(\gamma_{jr} \; ; R_j) \right\},$$

where $C\left(j, \tau(j)\right) = \dfrac{a_{\tau(j)} - \rho_{\tau(j)j} a_j}{\sqrt{1\text{-}\rho^2_{\tau(j)j}}}$ and $\gamma_{jr} = \dfrac{a_r - \rho_{rj} \cdot a_j}{\sqrt{1\text{-}\rho_{rj}{}^2}}$.

Q.E.D