School of Economic Sciences

Working Paper Series
WP 2013-4

Generalized Method of Truncated Moments Estimation of Censored Equation Systems

By

Faysal Fahs, Ron Mittelhammer, and Jonathan Yoder

March 2013

WASHINGTON STATE UNIVERSITY
World Class. Face to Face.

Generalized Method of Truncated Moments Estimation of Censored Equation Systems

Faysal Fahs, Ron Mittelhammer, and Jonathan Yoder *

ABSTRACT

We introduce a Generalized Method of Truncated Moments (GMTM) estimator as an alternative to a full maximum likelihood approach, and then use GMTM to define an estimator for a Tobit-type censored equation system. The GMTM implementation is based on a finite set of lower order moments derived from a fully parameterized multivariate Tobit model. The GMTM estimator is consistent, asymptotically normally distributed, near-asymptotically efficient, and more computationally tractable than maximum likelihood as model dimensionality increases. In addition, because only a truncated set of moments implied by a normally-distributed model are utilized in estimation, GMTM is more robust to distribution-specification error than the usual maximum likelihood estimator. Monte Carlo experiments based on various distributional assumptions are presented to compare the finite sample statistical properties of the GMTM approach to the ubiquitous Simulated Maximum Likelihood (SML) estimator, and suggest that the GMTM has similar performance to SML under normality and superior performance under non-normality, while also offering demonstrable computational advantages. Overall, the GMTM estimator is shown to be an empirically tractable and statistically attractive alternative method for estimating systems of censored regressions.

Keywords: Censored Equation Systems, Multivariate Tobit Model, GMM, Simulated Maximum Likelihood

JEL: Q13, L66

---

* Faysal Fahs is an Econometrician and AVP, PNC Financial Group, Inc.; Ron Mittelhammer (*corresponding author*) is a Regents Professor of Economic Sciences and Statistics, School of Economic Sciences, Washington State University; Jonathan Yoder is a Professor, School of Economic Sciences, Washington State University. The authors thank Scott Cardell, Rafic Fahs, Maher Hasan, and Jill McCluskey for helpful comments on evolving drafts of this paper.

GENERALIZED METHOD OF TRUNCATED MOMENTS ESTIMATION OF CENSORED EQUATION SYSTEMS

1. INTRODUCTION

Maximum likelihood (ML) is a commonly used estimation approach for univariate or multivariate censored regression models. In the case of the often-used Tobit-type censored model, ML requires the evaluation of a partially integrated multivariate normal density function, which is known to be computationally cumbersome, inaccurate, and increasingly intractable as disturbance distribution dimensionality increases (Lee, 1993). Similar multivariate integration challenges can arise for other choices of distributional assumptions.

Several methods have been proposed in attempts to improve estimation tractability of systems of censored regressions. Pudney (1989) estimates a system of Tobit equations by applying maximum likelihood to each equation marginally. Amemiya (1974) applied ML to a model that is based on only jointly positive dependent variable sample outcomes. While both techniques are consistent and numerically tractable, they are inefficient because they do not account for much of the information in the sample data. In particular, the former ignores cross-equation disturbance correlations and does not allow imposition of cross equation restrictions, while the latter omits information embodied in sample observations where one or more of the dependent variables is censored at its bound. Maddala (1977) modified Amemiya's (1974) procedure in a way that allowed all of the sample observations pertaining to the model to be utilized in estimation, but this procedure requires evaluation of partially integrated multivariate normal probability density functions of potentially high dimension, which is computationally burdensome, inaccurate, and/or completely intractable as the dimension of the model increases.

Hajivassiliou and Ruud (1994) proposed a Simulated Maximum Likelihood (SML) approach based on the Geweke-Hajivassiliou-Keane (GHK) simulator for overcoming the problem of high dimensional numerical integration underlying the choice probabilities in systems of censored equations. The SML method remains the most often-used method for solving ML problems that require integration of multivariate normal distribution for identifying likelihood function values. Advances in numerical computation methods for multivariate normal integrals showed promise for supporting alternative numerical approaches for calculating the likelihood contributions of sample observations (e.g., Genz (1992)), but, tractability, simplicity, computational burden and precision remain substantial constraints in the estimation of censored equation systems, especially for those of higher dimension.

This paper proposes a GMM approach for estimating fully parametric multivariate Tobit (MVT) models in which the characteristics of a multivariate probability distribution and its censoring process are summarized by a truncated set of its moments. This finite set of moments defines the estimating equation information for the estimator. We refer to this approach as *Generalized Method of Truncated Moments* (GMTM) estimation.

In our example implementation, we assume a multivariate Normal distribution and utilize moment conditions up to order $d = 3$, including univariate moment conditions together with pairwise cross-moments. Our Monte Carlo simulations suggest that this implementation of the GMTM estimator produces significant gains in tractability compared to a Simulated Maximum

Likelihood approach while also offering substantial efficiency in estimation as well as robustness with regard to alternative sampling distributions.

The remainder of the paper is organized as follows: Section 2 discusses the general conceptual underpinnings of the GMTM estimation concept. Section 3 reviews the basic structure of the MVT model in order to establish notation and define the particular estimation problem context in which the GMTM is applied. Section 4 describes the set of moment conditions used to define the estimating equation information in the implementation of the GMTM approach, and Section 5 motivates the explicit form of the GMTM estimator, including the application to the MVT model. Monte Carlo experiments are used to illustrate the computational and finite statistical properties of the GMM estimator in section 6, with the final section providing conclusions and implications of the proposed estimation method.

2. GENERAL GMTM CONCEPT

The GMTM approach is motivated by the classical "moment problem" in statistics, including the Hausdorf, Stieltjes, and Hamburger variants (e.g., Akhiezer(1965), Berg (1995), Fuglede (1965) and Shohat and Tamarki (1943)), and its relatively more recent variant, the "truncated moment problem" (e.g., Curto and Fialkow (1991, 2005) and Fialkow (2008)). A substantial literature has developed that determines various conditions under which a univariate or multivariate probability distribution is fully characterized by it moments, which is the classical "moments problem". For example, the normal family underlying the MVT model is such a distribution. A mathematical statement of the moment problem is essentially whether there exists a probability measure μ that satisfies the sequence of moment conditions

$$m_\alpha = \int_{x \in S} \boldsymbol{x}_1^{\alpha_1} \boldsymbol{x}_2^{\alpha_2} \cdots \boldsymbol{x}_J^{\alpha_J} \, d\mu, \;\; \alpha = \{\alpha_1, \alpha_2, \cdots, \alpha_J\}, \; \alpha_i \in \mathbb{Z}^{\geq} \;\; \forall i \qquad (2.1)$$

for a given sequence $\{m_\alpha\}$, where $\mathbb{Z}^{\geq}$ denotes the nonnegative integers and S denotes some support space.

A truncated moment problem of degree d addresses the types of measures μ that satisfy (2.1) under the additional restriction that $\sum_{j=1}^{J} \alpha_j \leq d$ for a given finite choice of $d \in \mathbb{Z}^{+}$. If a measure μ satisfies (2.1), it also satisfies the truncated moment conditions subset of a finite degree d, and the set of measures satisfying the truncated moment problem is generally larger than the singleton μ. Imposing only a subset of moment conditions of degree d implies less restrictions on the behavior of the underlying probability measures (see Stochel (2001) for a proof of these properties).

The GMTM approach adopts a truncated moments representation of the probability distribution underlying the specification of a parametric statistical model and its corresponding likelihood function and, as an alternative to ML estimation, estimates the model via GMM using the set of truncated moments. A goal of GMTM is to emulate, up to d-degree moment behavior, the probability distribution underlying the statistical model, with a degree sufficient to achieve near-asymptotic efficiency relative to the ML approach while also allowing for more robust behavior, via not constraining behavior beyond degree *d*, to accommodate uncertainty in the specification of the probability distribution. Moreover, to the extent that minimization of the GMM quadratic estimation objective function is a more tractable and/or stable optimization problem than ML, a goal of improving computational efficiency can also be achieved. Such is the case in the current application of GMTM to a system of censored regression functions of the MVT type.

Note that the use of higher order moments generally reduces the magnitude of the asymptotic covariance of the GMM estimator in the sense of positive definite symmetric matrix comparisons. However, the addition of higher order moments also enforces progressively more behavioral characteristics of the particular probability distribution specified in the statistical model, which progressively limits robustness of the method and adds to the complexity of the GMM estimation problem. In implementing GMTM, the degree of moments chosen is a balance between the efficiency achieved when the underlying probability distribution is correctly specified, the robustness achieved when that probability distribution is misspecified, and the computational complexity of the GMM estimation problem.

3. MULTIVARIATE TOBIT MODEL OVERVIEW

Consider the set of linear[1] regression equations

$$\boldsymbol{Y}_i^* = \begin{bmatrix} \mathbf{X}_{i1.} & & & \\ & \mathbf{X}_{i2.} & & \\ & & \ddots & \\ & & & \mathbf{X}_{iJ.} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_J \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{i1} \\ \boldsymbol{\varepsilon}_{i2} \\ \\ \boldsymbol{\varepsilon}_{iJ} \end{bmatrix} = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \qquad i = 1, \ldots, n, \qquad (3.1)$$

where $\boldsymbol{Y}_i^* \sim N(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma})$ is a $J \times 1$ vector of continuous latent variables that determine the probabilities of the observed outcomes of the dependent variables of the model (see equation 3.2 below), $\boldsymbol{\varepsilon}_i \sim iid\, N(\mathbf{0}, \boldsymbol{\Sigma})$ is a $J \times 1$ multivariate normal disturbance process, $\boldsymbol{\Sigma}$ is a

[1] All of the results can be generalized to nonlinear model specification by replacing the $\mathbf{X}_{ij.}\boldsymbol{\beta}_j$ terms with more general $\mathbf{g}(\mathbf{X}_{ij.}, \boldsymbol{\beta}_j)$ terms.

$J \times J$ covariance matrix, and *i* denotes the observation number. In (3.1), $\boldsymbol{X}_{ij.}$ denotes the i^{th} observation on a $1 \times K_j$ matrix of explanatory variables and $\boldsymbol{\beta}_j$ is the associated $K_j \times 1$ column vector of parameters, where the subscript *j* refers to the j^{th} member of the vector of J variables in $\boldsymbol{Y}_i^*$. Assuming that censoring is one-sided and occurs at the zero vector, observed variables Y_{ij} are characterized by

$$Y_{ij} = \begin{cases} Y_{ij}^* & \text{if } Y_{ij}^* > 0 \\ 0 & \text{otherwise} \end{cases} \qquad \text{for } i = 1,2,...,n, \text{ and } \quad j = 1,2,...,J \,. \tag{3.2}$$

Denoting outcomes of random variables by the use of lower case letters, for the i^{th} observation, the latent vector $\boldsymbol{y}_i^*$ contains J_d elements associated with discrete censored-at-zero outcomes $\boldsymbol{y}_{id} = 0$, and J_c elements corresponding to non-censored continuously measured outcomes, $\boldsymbol{y}_{ic} = \boldsymbol{y}_{ic}^* > 0$. Without loss of generality, we order the outcomes so that the $(J_d + J_c) \times 1$ vector $\boldsymbol{y}_i^*$ is defined as

$$\boldsymbol{y}_i^* = \begin{bmatrix} \boldsymbol{y}_{id}^* \\ \boldsymbol{y}_{ic}^* \end{bmatrix} \tag{3.3}$$

The joint probability density of the corresponding observed $\boldsymbol{y}_i = \begin{bmatrix} \boldsymbol{y}_{id} \\ \boldsymbol{y}_{ic} \end{bmatrix}$, which also represents the contribution of the i^{th} observation to the likelihood function, can then be represented as

$$L_i(\boldsymbol{\theta}; \boldsymbol{y}_i \;\; \boldsymbol{x}) = f(\boldsymbol{y}_i; \boldsymbol{x}, \boldsymbol{\theta}) = f_{ic}(\boldsymbol{y}_{ic}; \boldsymbol{x}, \boldsymbol{\theta}) \int_{-\infty}^{0} \cdots \int_{-\infty}^{0} f_{id}(\boldsymbol{y}_{id}^* \mid \boldsymbol{y}_{ic}; \boldsymbol{x}, \boldsymbol{\theta}) \, d\boldsymbol{y}_{id}^* \tag{3.4}$$

where $\boldsymbol{\theta} = [\boldsymbol{\beta}, \boldsymbol{\Sigma}]$ represents all of the model parameters, $f_{id}(\boldsymbol{y}_{id}^* \mid \boldsymbol{y}_{ic}; \boldsymbol{x}, \boldsymbol{\theta})$ represents the density of the random variables in the censored subset for the i^{th} observation, conditional on the variables in the non-censored set, and $f_{ic}\left(\boldsymbol{y}_{ic};\boldsymbol{x},\boldsymbol{\theta}\right)$ is the marginal probability density function for outcomes of the continuous random variables for the i^{th} observation.

The function $f_{ic}\left(\boldsymbol{y}_{ic};\boldsymbol{x},\boldsymbol{\theta}\right)$ contributing to the likelihood definition (3.4) is a J_c-dimensional multivariate normal probability density and its evaluation is straightforward. The contribution to the likelihood function represented by $\int_{-\infty}^{0}\cdots\int_{-\infty}^{0} f\left(\boldsymbol{y}_{id}^{*} \mid \boldsymbol{y}_{ic};\boldsymbol{x},\boldsymbol{\theta}\right)\boldsymbol{dy}_{id}^{*}$ is a J_d-dimensional integral of a conditional normal probability distribution, which is potentially difficult to evaluate and represents the source of the estimation difficulties noted in the introduction of the paper. In the next section we identify moments of the appropriate joint and conditional distributions of $\boldsymbol{Y}_i^*$, which are then used to construct a tractable GMTM estimator for the model parameters $\boldsymbol{\theta}$.

4. MOMENTS OF THE MVT MODEL

Univariate and bivariate moment conditions are presented in this section that together represent the estimating equation information used to identify and estimate the parameters of the MVT model via the truncated generalized method of moments approach. Univariate moments are defined exclusively in terms of one scalar random variable at a time. Bivariate moments are defined in terms of pairs of random variables, and include cross moment results in addition to truncation conditions defined in terms of random variable pairs. As noted in section 1, we are dealing with a finite collection of moments that, in the context of the classical “moment problem” in statistics, represents a truncated moment characterization of the underlying multivariate normal probably distribution. In the current implementation, we focus on moments up to third order.

4.1 *Univariate moment conditions*

The first and second order conditional-on-positivity univariate moments of a normal random variable have been derived by Amemiya (1973), and are given by

$$E\left(\boldsymbol{Y}_{.j} \mid \boldsymbol{Y}_{.j} > \mathbf{0}\right) = \boldsymbol{U}_{.j} + \sigma_j \left(\boldsymbol{\phi}_j \oslash \boldsymbol{\Phi}_j\right) \tag{4.1.1}$$

$$E\left(\boldsymbol{Y}_{.j}^2 \mid \boldsymbol{Y}_{.j} > \mathbf{0}\right) = \boldsymbol{U}_{.j} \odot E\left(\boldsymbol{Y}_{.j} \mid \boldsymbol{Y}_{.j} > \mathbf{0}\right) + \sigma_j^2 \tag{4.1.2}$$

where $\boldsymbol{Y}_{.j}$ denotes *n* observations on the *j*th dependent variable, $\mathbf{U}_{.j} \equiv \mathbf{X}_{.j}\boldsymbol{\beta}_j$ denotes *n* observations on the systematic component associated with the *j*th dependent variable, $\mathbf{X}_{.j}$ is an $n \times K_j$ matrix of vertically concatenated explanatory $\mathbf{X}_{ij.}$ vectors for $i = 1,...,n$, σ_j^2 is the variance associated with the *j*th variable, $\boldsymbol{\phi}_j \equiv \boldsymbol{\phi}\left(\frac{\mathbf{U}_{.j}}{\sigma_j}\right)$ denotes *n* observations on the standard normal probability density function (PDF) evaluated at the *n* arguments $\frac{\mathbf{U}_{.j}}{\sigma_j}$, $\boldsymbol{\Phi}_j \equiv \boldsymbol{\Phi}\left(\frac{\mathbf{U}_{.j}}{\sigma_j}\right)$ denotes *n* observation on the cumulative distribution function (CDF) of the standard normal distribution, again evaluated at $\frac{\mathbf{U}_{.j}}{\sigma_j}$, and $\oslash$ and $\odot$ are the Hadamard (i.e., elementwise) division and multiplication operators, respectively. Equations (4.1.1) and (4.1.2) represent the first and second-order conditional univariate moments for positive observations $\boldsymbol{Y}_{.j} > \mathbf{0}$.

Using all sample observations (including censored observations) in $\boldsymbol{Y}_{j.}$, the two *unconditional* univariate moments for $\boldsymbol{Y}_{j.}$, with orders corresponding to (4.1.1) and (4.1.2), are given by (Heckman, 1976*b*):

$$E\left(\boldsymbol{Y}_{.j}\right)=\boldsymbol{U}_{.j}\odot\boldsymbol{\Phi}_{j}+\sigma_{j}\boldsymbol{\phi}_{j} \tag{4.1.3}$$

$$E\left(\boldsymbol{Y}_{.j}^{2}\right)=\boldsymbol{U}_{.j}\odot E\left(\boldsymbol{Y}_{.j}\right)+\sigma_{j}^{2}\boldsymbol{\Phi}_{j}\,. \tag{4.1.4}$$

We define an additional set of moment conditions by defining a binary variable, analogous to its use in the univariate Probit case, as

$$\boldsymbol{Y}_{.j}^{Bin}=\begin{cases}1 & if \quad \boldsymbol{Y}_{.j}>\boldsymbol{0}\\ 0 & \text{otherwise}\end{cases}$$

The binary univariate moment condition can be represented as

$$\mathrm{E}(\boldsymbol{Y}_{.j}^{Bin})=\boldsymbol{\Phi}_{j} \tag{4.1.5}$$

where $\boldsymbol{\Phi}_{j}$ is defined as above.

Collecting all of the moment conditions (4.1.1)–(4.1.5) for all *n* observations, we define the following relationships relating $\boldsymbol{U}_{.j}$ and $\boldsymbol{Y}_{.j}$, together with a collection of disturbances $\boldsymbol{\xi}_{j}^{(i)}$ having zero expectations, as:

$$\begin{aligned}
\boldsymbol{Y}_{.j+}&=\boldsymbol{U}_{.j+}+\sigma_{j}\left(\boldsymbol{\phi}_{j+}\oslash\boldsymbol{\Phi}_{j+}\right)+\boldsymbol{\xi}_{j}^{(1)}\\
\boldsymbol{Y}_{.j+}^{2}&=\boldsymbol{U}_{.j+}\odot E(\boldsymbol{Y}_{.j+})+\sigma_{j}^{2}+\boldsymbol{\xi}_{j}^{(2)}\\
\boldsymbol{Y}_{.j}&=\boldsymbol{U}_{.j}\odot\boldsymbol{\Phi}_{j}+\sigma_{j}\boldsymbol{\phi}_{j}+\boldsymbol{\xi}_{j}^{(3)}\\
\boldsymbol{Y}_{.j}^{2}&=\boldsymbol{U}_{.j}\odot E\left(\boldsymbol{Y}_{.j}\right)+\sigma_{j}^{2}\Phi_{j}+\boldsymbol{\xi}_{j}^{(4)}\\
\boldsymbol{Y}_{.j}^{Bin}&=\Phi_{j}+\boldsymbol{\xi}_{j}^{(5)}
\end{aligned} \tag{4.1.6}$$

where $\mathbf{Y}_{.j+}$ denotes the set of positive-valued observations relating to the j^{th} dependent variable, $\boldsymbol{U}_{.j+}$ denotes observations on the systematic components that correspond to the positive-valued outcomes, and $E\left(\boldsymbol{Y}_{.j+}\right)$ is shorthand notation $E\left(\boldsymbol{Y}_{.j+}\right) \equiv E\left(\boldsymbol{Y}_{.j} \mid \boldsymbol{Y}_{.j} > \mathbf{0}\right)$. In this formulation, $\boldsymbol{\Phi}_j$ and $\boldsymbol{\phi}_j$ are vectors of the cumulative distribution function and density function values of the standard normal distribution, evaluated at the vector $\left(\frac{1}{\sigma_j}\right)\boldsymbol{U}_{.j}$, and $\boldsymbol{\Phi}_{j+}$ and $\boldsymbol{\phi}_{j+}$ are the subsets of those vectors corresponding to the positive valued observations $\boldsymbol{Y}_{.j+}$. Based on (4.1.6), $(5K_j \times 1)$ orthogonality conditions can be defined based on the preceding univariate moments:

$$E\left[\boldsymbol{h}_j^{Uni}\left(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\theta}\right)\right] \equiv E\begin{bmatrix} \left[\boldsymbol{X}'_{.j+}\left(\boldsymbol{Y}_{.j+} - \boldsymbol{U}_{.j+} - \sigma_j\left(\boldsymbol{\phi}_{.j+} \oslash \boldsymbol{\Phi}_{.j+}\right)\right)\right] \\ \left[\boldsymbol{X}'_{.j+}\left(\boldsymbol{Y}^2_{.j+} - \boldsymbol{U}_{.j+} \odot E\left(\boldsymbol{Y}_{.j+}\right) - \sigma_j^{\,2}\right)\right] \\ \left[\boldsymbol{X}'_{.j}\left(\boldsymbol{Y}_{.j} - \boldsymbol{U}_{.j} \odot \boldsymbol{\Phi}_j - \sigma_j \boldsymbol{\phi}_j\right)\right] \\ \left[\boldsymbol{X}'_{.j}\left(\boldsymbol{Y}_{.j}^{\,2} - \boldsymbol{U}_{.j} \odot E\left(\boldsymbol{Y}_{.j}\right) - \sigma_j^{\,2}\boldsymbol{\Phi}_j\right)\right] \\ \left[\boldsymbol{X}'_{.j}\left(\boldsymbol{Y}^{Bin}_{.j} - \boldsymbol{\Phi}_j\right)\right] \end{bmatrix} = \mathbf{0} \tag{4.1.7}$$

where $\boldsymbol{X}_{.j+}$ denotes the $n_{j+} \times K_j$ matrix of observations on the explanatory variables that correspond to the n_{j+} positive-valued outcomes for the j^{th} dependent variable. The sample analog of the population moments displayed in (4.1.7) is

$$\boldsymbol{h}_j^{Uni}\left(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\theta}\right)=\begin{bmatrix} \left(\boldsymbol{x}'_{.j+}/n_{j+}\right)\left(\boldsymbol{y}_{.j+} - \boldsymbol{u}_{.j+} - \sigma_j \odot \left(\boldsymbol{\phi}^o_{ij+} \oslash \boldsymbol{\Phi}^o_{ij+}\right)\right) \\ \left(\boldsymbol{x}'_{.j+}/n_{j+}\right)\left(\boldsymbol{y}^2_{.j+} - \boldsymbol{u}_{.j+} \odot E\left(\boldsymbol{y}_{.j+}\right) - \sigma_j^{\,2}\right) \\ \left(\boldsymbol{x}'_{.j}/n\right)\left(\boldsymbol{y}_{.j} - \boldsymbol{u}_{.j} \odot \boldsymbol{\Phi}^o_j - \sigma_j \boldsymbol{\phi}^o_j\right) \\ \left(\boldsymbol{x}'_{.j}/n\right)\left(\boldsymbol{y}_{.j}^{\,2} - \boldsymbol{u}_{.j} \odot E\left(\boldsymbol{y}_{.j}\right) - \sigma_j^{\,2}\boldsymbol{\Phi}^o_j\right) \\ \left(\boldsymbol{x}'_{.j}/n\right)\left(\boldsymbol{y}^{Bin}_{.j} - \boldsymbol{\Phi}^o_j\right) \end{bmatrix} \quad (4.1.8)$$

where $E\left(\boldsymbol{y}_{.j+}\right)$ denotes $E\left(\boldsymbol{Y}_{.j+}\right)$ evaluated at sample outcomes for $\boldsymbol{y}_{.j}$ and $\boldsymbol{x}_{.j}$ for specified values of $\boldsymbol{\theta}$, and more generally the change from capital letters to lower case letters again indicates that we are evaluating terms at sample values, with superscript “*o*” on $\boldsymbol{\phi}$ and $\boldsymbol{\Phi}$ also denoting evaluation at observed sample values.

Collecting the moment conditions across all J dependent variables in the censored system, the dimension of the full set of univariate moment conditions, $\boldsymbol{h}^{Uni}\left(\mathbf{Y},\mathbf{X},\boldsymbol{\theta}\right) \equiv \left[\boldsymbol{h}_1^{Uni}\left(\mathbf{Y},\mathbf{X},\boldsymbol{\theta}\right)' \cdots \boldsymbol{h}_J^{Uni}\left(\mathbf{Y},\mathbf{X},\boldsymbol{\theta}\right)'\right]'$, is equal to $\left(5\sum_{j=1}^{J} K_j \times 1\right)$.

4.2 *Bivariate moment conditions*

Bivariate moment information is particularly useful in identifying and estimating the parameters involved in the cross-equation disturbance covariance structure. Fortunately, the numerical calculation of bivariate moment information is accurate and computationally fast.

Tallis (1961) derived the first two moments and the moment generating function of the truncated multivariate normal distribution. Extending Tallis, Fahs and Mittelhammer (2001) derived the first two moments of the truncated bivariate normal distribution based on a more familiar standardized variable parameterization of the model that is more convenient for use within the current GMM framework. We additionally utilize third order moments along with cross-moment results in the specification of the GMTM estimator.

The MVT model contains J dependent variables, so that there are $J(J-1)/2$ unique bivariate pairs. For example, in a five dimensional model, there are ten unique bivariate pairs given by (y_{ji}, y_{ki}), j *and* $k \in \{1,2,3,4,5\}$, $j<k$. Nine different moment conditions through the third order for each unique pair (y_{ji}, y_{ki}) can be defined as follows:

First order: $E(y_{ji} \mid y_{ji} \geq 0, y_{ki} \geq 0)$ and $E(y_{ki} \mid y_{ji} \geq 0, y_{ki} \geq 0)$

Second order: $E(y_{ji}^2 \mid y_{ji} \geq 0, y_{ki} \geq 0)$ and $E(y_{ki}^2 \mid y_{ji} \geq 0, y_{ki} \geq 0)$

Third order: $E(y_{ji}^3 \mid y_{ji} \geq 0, y_{ki} \geq 0)$ and $E(y_{ki}^3 \mid y_{ji} \geq 0, y_{ki} \geq 0)$

Cross moments: $E(y_{ji}y_{ki} \mid y_{ji} \geq 0, y_{ki} \geq 0)$, $E(y_{ji}y_{ki}^2 \mid y_{ji} \geq 0, y_{ki} \geq 0)$, and

$E(y_{ji}^2 y_{ki} \mid y_{ji} \geq 0, y_{ki} \geq 0)$.

Definitions for all of the expectation functions referred to above are provided in the Appendix.

A set of orthogonality conditions $\boldsymbol{h}^{Biv}(\mathbf{Y},\mathbf{X},\boldsymbol{\theta}) \equiv \left[\boldsymbol{h}_1^{Biv}(\mathbf{Y},\mathbf{X},\boldsymbol{\theta})' \cdots \boldsymbol{h}_J^{Biv}(\mathbf{Y},\mathbf{X},\boldsymbol{\theta})'\right]'$ defined on the basis of the moment conditions listed above can be specified as

$$E\left[\boldsymbol{h}_j^{Biv}(\boldsymbol{Y},\boldsymbol{X},\boldsymbol{\theta})\right] = E\, vec\left(\begin{bmatrix} \boldsymbol{X}_{.j}'\left(\boldsymbol{Y}_{.j} - E\left(\boldsymbol{Y}_{.j} \mid \boldsymbol{Y}_{.j} \geq 0, \boldsymbol{Y}_{.k} \geq 0\right)\right) \\ \boldsymbol{X}_{.j}'\left(\boldsymbol{Y}_{.j}^2 - E\left(\boldsymbol{Y}_{.j}^2 \mid \boldsymbol{Y}_{.j} \geq 0, \boldsymbol{Y}_{.k} \geq 0\right)\right) \\ \boldsymbol{X}_{.j}'\left(\boldsymbol{Y}_{.j}^3 - E\left(\boldsymbol{Y}_{.j}^3 \mid \boldsymbol{Y}_{.j} \geq 0, \boldsymbol{Y}_{.k} \geq 0\right)\right) \\ \boldsymbol{X}_{.j}'\left(\boldsymbol{Y}_{.j} \odot \boldsymbol{Y}_{.k} - E\left(\boldsymbol{Y}_{.j} \odot \boldsymbol{Y}_{.k} \mid \boldsymbol{Y}_{.j} \geq 0, \boldsymbol{Y}_{.k} \geq 0\right)\right) \\ \boldsymbol{X}_{.j}'\left(\boldsymbol{Y}_{.j}^2 \odot \boldsymbol{Y}_{.k} - E\left(\boldsymbol{Y}_{.j}^2 \odot \boldsymbol{Y}_{.k} \mid \boldsymbol{Y}_{.j} \geq 0, \boldsymbol{Y}_{.k} \geq 0\right)\right) \\ \boldsymbol{X}_{.j}'\left(\boldsymbol{Y}_{.j} \odot \boldsymbol{Y}_{.k}^2 - E\left(\boldsymbol{Y}_{.j} \odot \boldsymbol{Y}_{.k}^2 \mid \boldsymbol{Y}_{.j} \geq 0, \boldsymbol{Y}_{.k} \geq 0\right)\right) \end{bmatrix}, k \neq j\right) = \mathbf{0} \qquad (4.2.1)$$

for $j=1,\dots,J$. There are a maximum of $\left(6(J-1)\sum_{j=1}^{J}K_j\right)$ different conditions represented by $\boldsymbol{h}^{Biv}(\mathbf{Y},\mathbf{X},\boldsymbol{\theta})$ given that there are no redundant vectors among the columns of $X_{j.}, j=1,\dots,J$. The number of effective orthogonality conditions would be reduced by the number of redundancies[2].

The sample analog of the orthogonality conditions (4.2.1) is given by

$$\boldsymbol{h}_j^{Biv}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\theta}) = vec\left(\begin{bmatrix} (\boldsymbol{x}'_{.j}/n_{jk})(\boldsymbol{y}_{.j}-E(\boldsymbol{y}_{.j}\mid \boldsymbol{y}_{.j}\geq 0,\boldsymbol{y}_{.k}\geq 0)) \\ (\boldsymbol{x}'_{.j}/n_{jk})(\boldsymbol{y}_{.j}^2-E(\boldsymbol{y}_{.j}^2\mid \boldsymbol{y}_{.j}\geq 0,\boldsymbol{y}_{.k}\geq 0)) \\ (\boldsymbol{x}'_{.j}/n_{jk})(\boldsymbol{y}_{.j}^3-E(\boldsymbol{y}_{.j}^3\mid \boldsymbol{y}_{.j}\geq 0,\boldsymbol{y}_{.k}\geq 0)) \\ (\boldsymbol{x}'_{.j}/n_{jk})(\boldsymbol{y}_{.j}\odot\boldsymbol{y}_{.k}-E(\boldsymbol{y}_{.j}\odot\boldsymbol{y}_{.k}\mid \boldsymbol{y}_{.j}\geq 0,\boldsymbol{y}_{.k}\geq 0)) \\ (\boldsymbol{x}'_{.j}/n_{jk})(\boldsymbol{y}_{.j}^2\odot\boldsymbol{y}_{.k}-E(\boldsymbol{y}_{.j}^2\odot\boldsymbol{y}_{.k}\mid \boldsymbol{y}_{.j}\geq 0,\boldsymbol{y}_{.k}\geq 0)) \\ (\boldsymbol{x}'_{.j}/n_{jk})(\boldsymbol{y}_{.j}\odot\boldsymbol{y}_{.k}^2-E(\boldsymbol{y}_{.j}\odot\boldsymbol{y}_{.k}^2\mid \boldsymbol{y}_{.j}\geq 0,\boldsymbol{y}_{.k}\geq 0)) \end{bmatrix}, k\neq j\right) = \mathbf{0} \qquad (4.2.2)$$

where n_{jk} denotes the number of sample observations involved in the $(j,k)^{th}$ pair.

The total number of estimating equations in the vectors $\boldsymbol{h}^{Uni}(\mathbf{Y},\mathbf{X},\boldsymbol{\theta})$ and $\boldsymbol{h}^{Biv}(\mathbf{Y},\mathbf{X},\boldsymbol{\theta})$ equals $(6J-1)\sum_{j=1}^{J}K_j$, which will overdetermine the number of unknown parameters in the model, $\left(J(J-1)/2+\sum_{j=1}^{J}K_j\right)$, in all problems of practical interest. Consequently, there is no unique parameter vector $\boldsymbol{\theta}$ that solves the estimating equations. A GMM estimation approach is developed next.

[2] For one extreme example, if $\boldsymbol{X}'_{.j} = \boldsymbol{X}'_{.k}$, then the vector $\boldsymbol{X}'_{.j}\left(\boldsymbol{Y}_{.j}\odot\boldsymbol{Y}_{.k}-E(\boldsymbol{Y}_{.j}\odot\boldsymbol{Y}_{.k}\mid \boldsymbol{Y}_{.j}\geq 0,\boldsymbol{Y}_{.k}\geq 0)\right)$ is in fact identical to the vector $\boldsymbol{X}'_{.k}\left(\boldsymbol{Y}_{.k}\odot\boldsymbol{Y}_{.j}-E(\boldsymbol{Y}_{.k}\odot\boldsymbol{Y}_{.j}\mid \boldsymbol{Y}_{.k}\geq 0,\boldsymbol{Y}_{.j}\geq 0)\right)$.

5. GMM ESTIMATION OF THE MVT SYSTEM

The full set of population moment conditions identified in section 4 is represented by the vertical concatenation of (4.1.7) and (4.2.1), as

$$E\left(\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{X},\boldsymbol{\theta})\right)=E\begin{bmatrix}\boldsymbol{h}^{Uni}(\boldsymbol{Y},\boldsymbol{X},\boldsymbol{\theta})\\ \boldsymbol{h}^{Bin}(\boldsymbol{Y},\boldsymbol{X},\boldsymbol{\theta})\end{bmatrix}=\mathbf{0}, \tag{5.1}$$

and $\boldsymbol{h}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\theta})=\begin{bmatrix}\boldsymbol{h}^{Uni}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\theta})\\ \boldsymbol{h}^{Biv}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\theta})\end{bmatrix}$ is the sample moment analog to (5.1). The GMM parameter vector is chosen such that the sample moment conditions are as close to the zero vector as possible in terms of weighted Euclidean distance, and the GMM objective function is

$$\min_{\boldsymbol{\theta}}\left[Q(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\theta})\right]=\min_{\boldsymbol{\theta}}\left[\ \boldsymbol{h}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\theta})'\,\boldsymbol{W}\ \boldsymbol{h}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\theta})\right] \tag{5.2}$$

where ***W*** is a conformable positive definite symmetric weight matrix. The first order conditions for the minimization problem in (5.2) are given by

$$\frac{\partial Q(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}=2\left[\frac{\partial\boldsymbol{h}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]'\boldsymbol{W}\,\boldsymbol{h}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\theta})=\mathbf{0}\,. \tag{5.3}$$

The choice of ***W*** that results in the asymptotically most efficient estimator within the class of GMM estimators based on the moment equations (5.1) is the inverse of the covariance of the estimating equations, $\boldsymbol{W}^*=\left[\mathbf{Cov}\left(\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{X},\boldsymbol{\theta})\right)\right]^{-1}=E\left(\boldsymbol{hh}'\right)^{-1}$ (Hansen, 1982; Andrews, 1999). The optimal weight matrix $\boldsymbol{W}^*$ is unknown, and a consistent estimator, $\hat{\boldsymbol{W}}_n$, is needed to ensure consistency and asymptotic efficiency of the GMM estimator. In practice, this is generally obtained by first setting ***W*** = ***I*** and calculating $\hat{\boldsymbol{\theta}}(\boldsymbol{I})$ in (5.2) and, in the process, estimating the covariance matrix of ***h***. Then in a second step, the sample estimator of the optimal weighting matrix $\hat{\boldsymbol{W}}_n$ is substituted into (5.2) leading to the estimated optimal GMM defined by $\hat{\boldsymbol{\theta}}_{GMM}(\hat{\boldsymbol{W}}_n)$. The estimated optimal GMM estimator will be consistent, asymptotically normal and asymptotically efficient estimator in the sense of making asymptotically efficient use of the given moment information used in estimation.

The estimation of the model parameters based on the GMTM method is summarized as follows:

1. Define all conditional (truncated) and unconditional first and second order univariate sample moments (4.1.8), and all possible combinations of the first, second and third order conditional (truncated) bivariate sample moments, including cross moments of the form (4.2.2).
2. Estimate the optimal weight matrix $\boldsymbol{W}^*$ based on a consistent estimate of the covariance matrix of the moments ***h***.
3. Minimize the GMM criterion function (5.2) based on the optimal weighting matrix $\hat{\boldsymbol{W}}_n$.

Given that the estimation objective function is nonlinear in parameters, starting values are required for any numerical minimization algorithm. We suggest using the univariate and bivariate Tobit estimators for this purpose, implemented as follows. First, *J* univariate Tobit models are estimated, one for each dependent variable, to provide estimates $\hat{\boldsymbol{\beta}}_j$ and $\hat{\sigma}_j^2$ for each $j \in J$. Second, a bivariate Tobit model is estimated for every distinct pair of dependent variables to provide alternative estimates of $\hat{\boldsymbol{\beta}}_j$ and for the $J(J-1)/2$ unique covariance parameters. Multiple estimates of parameters obtained through the use of both the univariate and bivariate Tobit model estimations are averaged, and then the averaged values of the estimates are used as starting values for solving the GMM estimation problem (5.2).

6. MONTE CARLO EXPERIMENTS

In this section we perform Monte Carlo experiments to compare the parameter estimates of the GMTM approach to SML estimates (Hajivassiliou and Ruud 1994). We begin with a two-equation model and compare estimation performance of the two approaches under several different disturbance distributions, and we also test moment equation validity under the different distributions. In a second experiment we expand the MVT system to a five equation model and compare the GMM to the SML under normality, the latter being the asymptotically optimal estimation method under the multivariate normal sampling process.

For the minimization of equation (5.2), we use the Nelder-Meade polytope direct-search method of optimization. It is robust to non-differentiability (the method requires only that the function being minimized is continuous) and is useful for functions whose derivatives cannot be calculated or approximated easily, or at all. A convergence criterion of 0.00001 was used for the difference between the maximum and the minimum objective function associated with the vertices of the Nelder-Meade simplex. We utilized the GAUSS™ Mathematical and Statistical System to program the GMTM method. The SML estimates were obtained following the now standard approach outlined in Hajivassiliou and Ruud (1994).

6.1 *Sampling models*

We begin generating outcomes of the latent variable in equation (3.1) by sampling the columns of $\boldsymbol{X}$ from uniform distributions having support (-5, 5). The two-equation model is specified as

$$\begin{bmatrix}\mathbf{Y}_1^* \\ \mathbf{Y}_2^*\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1\boldsymbol{\beta}_1 \\ \mathbf{X}_2\boldsymbol{\beta}_2\end{bmatrix} + \begin{bmatrix}\boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2\end{bmatrix},\ \boldsymbol{\beta}_1 = \begin{bmatrix}\beta_{11} \\ \beta_{12}\end{bmatrix} = \begin{bmatrix}.1 \\ .2\end{bmatrix},\ \boldsymbol{\beta}_2 = \begin{bmatrix}\beta_{21} \\ \beta_{22}\end{bmatrix} = \begin{bmatrix}.3 \\ .4\end{bmatrix}$$

$$\begin{bmatrix}\boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2\end{bmatrix} \sim \left(\mathbf{0}, \boldsymbol{S}^{-1}\mathbf{R}\boldsymbol{S}^{-1}\right),\ \mathbf{R} = \begin{bmatrix}1 & .2 \\ .2 & 1\end{bmatrix}, \boldsymbol{S} = \begin{bmatrix}1 & 0 \\ 0 & 2\end{bmatrix}. \tag{6.1.1}$$

where $\boldsymbol{R}$ is the correlation matrix and $\boldsymbol{S}$ is the diagonal matrix of standard deviations**.** The various disturbance distributions are discussed in section 6.2 below. The five-equation model is

$$\begin{bmatrix}\mathbf{Y}_1^* \\ \mathbf{Y}_2^* \\ \mathbf{Y}_3^* \\ \mathbf{Y}_4^* \\ \mathbf{Y}_5^*\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1\boldsymbol{\beta}_1 \\ \mathbf{X}_2\boldsymbol{\beta}_2 \\ \mathbf{X}_3\boldsymbol{\beta}_3 \\ \mathbf{X}_4\boldsymbol{\beta}_4 \\ \mathbf{X}_5\boldsymbol{\beta}_5\end{bmatrix} + \begin{bmatrix}\boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \boldsymbol{\varepsilon}_3 \\ \boldsymbol{\varepsilon}_4 \\ \boldsymbol{\varepsilon}_5\end{bmatrix},\ \boldsymbol{\beta}_1 = \begin{bmatrix}.1 \\ .2\end{bmatrix}, \boldsymbol{\beta}_2 = \begin{bmatrix}.3 \\ .4\end{bmatrix}, \boldsymbol{\beta}_3 = \begin{bmatrix}.5 \\ .6\end{bmatrix}, \boldsymbol{\beta}_4 = \begin{bmatrix}.7 \\ .8\end{bmatrix}, \boldsymbol{\beta}_5 = \begin{bmatrix}.5 \\ .2\end{bmatrix},$$

$$\begin{bmatrix}\boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \boldsymbol{\varepsilon}_3 \\ \boldsymbol{\varepsilon}_4 \\ \boldsymbol{\varepsilon}_5\end{bmatrix} \sim \left(\mathbf{0}, \boldsymbol{S}^{-1}\boldsymbol{R}\boldsymbol{S}^{-1}\right),\ \boldsymbol{R} = \begin{bmatrix}1 & .1 & .2 & .3 & .4 \\ .1 & 1 & .6 & .4 & .25 \\ .2 & .6 & 1 & .11 & .12 \\ .3 & .4 & .11 & 1 & .15 \\ .4 & .25 & .12 & .15 & 1\end{bmatrix}, \boldsymbol{S} = \text{diag}\begin{pmatrix}1 \\ 2 \\ 4 \\ 3 \\ 4\end{pmatrix}. \tag{6.1.2}$$

The correlation matrix, $\boldsymbol{R}$, and the standard deviations $\boldsymbol{S}$ in (6.1.2) were chosen so as to produce an appreciable degree of disturbance variability as well as to produce a range of covariances between latent variables. The preceding two specifications are used to generate the latent variables and the data for the MVT model presented in equation (3.2).

6.2 *Two-equation Tobit model performance*

Table 1 summarizes the performance of the GMM and SML estimators when sampling from a multivariate normal distribution. With a large sample size of 1,000 observations SML outperforms GMTM only slightly in term of root mean square error (RMSE).[3] The fact that SML would be at least nominally superior is not surprising given that the Tobit SML estimator is asymptotically efficient, achieving the Cramér-Rao lower bound, when the data generating process is actually normally distributed. For equation 1 (containing parameters β_{11} and β_{12}) the censoring percentage is 23%, while for equation 2 (containing parameters β_{21} and β_{22}) censoring occurs 20% of the time. We discuss later how the censoring rate affects results.

Table 1: Comparison between GMTM and SML, Two-Equation Model, Sample Size = 1000, Repetitions = 1000, Under the Assumption of Normality

Parameters	True Values	GMTM (MEAN)	GMTM (RMSE)	SML (MEAN)	SML (RMSE)	T-Test Values (RMSE)	% Censored Obs per Equation
β_{11}	0.1000	0.1054	0.0475	0.1032	0.0462	0.7888	23%
β_{12}	0.2000	0.1998	0.0310	0.1991	0.0299	0.7280	
β_{21}	0.3000	0.2940	0.0742	0.2957	0.0731	0.9775	20%
β_{22}	0.4000	0.3996	0.0482	0.4008	0.0442	1.1747	
ρ	0.2000	0.2088	0.0088	0.2013	0.0083	0.8375	
Time to Convergence (Minutes)		**140**		**310**			

A t-test was performed to assess statistical significance of differences in the RMSEs under GMTM and SML. The null hypothesis for the RMSEs is H_0: mean($RMSE_{GMTM}$) = mean($RMSE_{SML}$). The results in Table 1 show that under the assumption of normality we fail to reject the null

[3] $\text{RMSE} = \sqrt{m^{-1}\sum_{i=1}^{m}\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_i\right)^2}$, where $\boldsymbol{\theta}$ is the vector of true parameter values and $\hat{\boldsymbol{\theta}}_i$ is vector of estimated parameters, with m being the MC sample size of 1000 for all simulations.

hypotheses of equality for the RMSEs, supporting the contention that GMTM is nearly as asymptotically efficient as the SML estimator for reasonably large sample sizes. Moreover, GMTM exhibited two advantages over SML, tractability and speed of convergence. The convergence time shows that GMM converged, on average, 55% faster than SML.

Table 2: Comparison between GMTM and SML, Two-Equation Model, Sample Size = 100, Repetitions = 1000, Under the Assumption of Normality

Parameters	True Values	GMTM (MEAN)	GMTM (RMSE)	SML (MEAN)	SML (RMSE)	T-Test Values (RMSE)	% Censored Obs per Equation
β_{11}	0.1000	0.0922	0.1016	0.0952	0.1670	1.9930	23%
β_{12}	0.2000	0.2012	0.0629	0.1974	0.1119	2.0669	
β_{21}	0.3000	0.2948	0.1767	0.3042	0.2387	1.9942	20%
β_{22}	0.4000	0.4125	0.0999	0.4139	0.1499	2.0206	
ρ	0.2000	0.2094	0.0130	0.1870	0.0294	1.9984	
Time to Convergence (Minutes)		**97**		**210**			

Continuing with the assumption of normality, Table 2 shows that decreasing the sample size by a factor of 10 (to *n=100*) causes the RMSEs for both approaches to increase. However, the magnitudes of the RMSEs from GMTM are now *smaller* than the magnitudes of the RMSEs from SML, suggesting that GMTM is an improvement over SML under the conditions of these simulations. The t-test for the RMSEs confirms the statistical significance of the RMSE improvement afforded by GMTM. Again the censoring percentage for each equation is appreciable, being 23% and 20% respectively, and GMTM converges 54% faster than SML on average. In summary, under the assumption of normality and under an appreciable degree of censoring, the above results suggest the following comparison of GMTM and SML behavior in the preceding Two-Equation Tobit model:

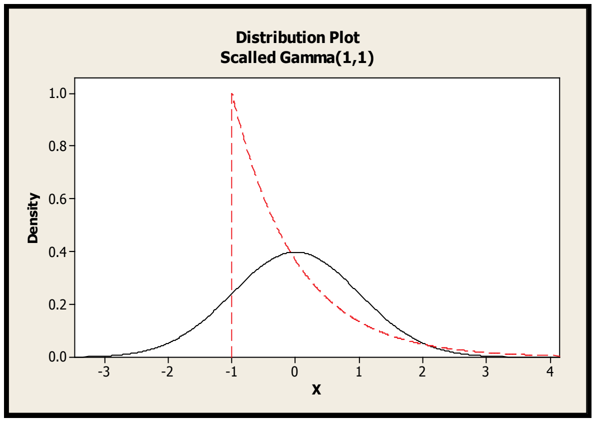
1. For the large sample (*n* = 1000), GMTM is nearly as efficient as SML.

2. GMTM is computationally tractable and converges faster for both the large (n = 1000) and small (n = 100) sample sizes.
3. GMTM is more efficient for the small sample size compared to SML.

We also compare GMTM and SML behavior under three types of Gamma distributions with varying degrees of skewness. Figure 1 provides graphical representations of these distributions. Each distribution is centered and scaled to have a mean of zero and variance of one.

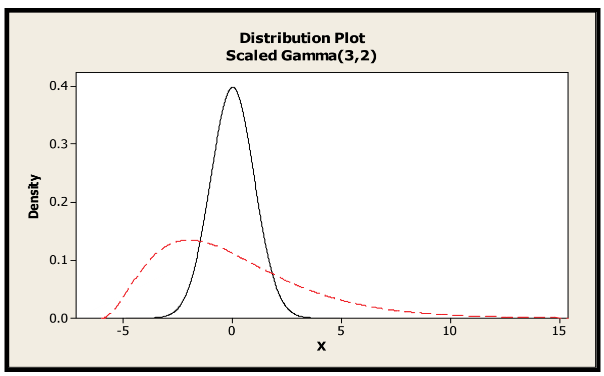
Tables 3, 4, and 5 provide comparisons of GMTM and SML sampling behavior based on centered and scaled Gamma(1,1), Gamma (3,2), and Gamma (4,3) distributions, respectively, with a sample size of 1000 observations. For all of these cases, GMTM is seen to be an improvement over the SML in terms of RMSE. The RMSEs magnitudes of GMTM are uniformly smaller than those of SML, and the t-tests of RMSE equality also indicate that there are statistically significant differences between the average simulated RMSEs of the two approaches. This finding underscores that the Tobit SML estimator becomes inefficient as the data generating process deviates from the normality assumption, and suggests further that GMTM offers a degree of robustness against the effects of distributional misspecification, derived from truncating the required moment behavior implied by the multivariate normal distribution used in this implementation.

Figure 1: Graphical representations of centered and scaled Gamma distributions



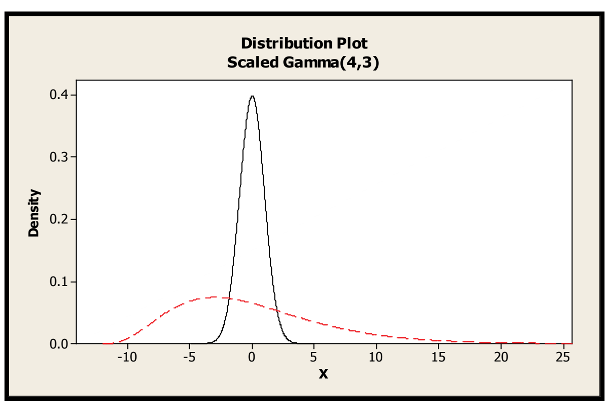
***Dashed line*: Centered and Scaled Gamma (1,1) distribution**

***Solid line* : N(0,1) distribution**



***Dashed line*: Centered and Scaled Gamma (3,2) distribution**

***Solid line* : N(0,1) distribution**



***Dashed line*: Centered and Scaled Gamma (4,3) distribution**

***Solid line* : N(0,1) distribution**

Table 3: Comparison between GMTM and SML, Two-Equation Model, Sample Size = 1000, Repetitions = 1000, Under the Assumption of Gamma (1,1)

Parameters	**True Values**	**GMTM (MEAN)**	**GMTM (RMSE)**	**SML (MEAN)**	**SML (RMSE)**	**T-Test Values (RMSE)**	**% Censored Obs per Equation**
β_{11}	0.1000	0.0867	0.0156	0.1238	0.0447	2.0772	56.6%
β_{12}	0.2000	0.2117	0.0182	0.2206	0.0825	3.1256	
β_{21}	0.3000	0.2768	0.0268	0.2555	0.3697	2.8754	49.5%
β_{22}	0.4000	0.447	0.0109	0.5011	0.0161	1.9824	
ρ	0.2000	0.1983	0.0102	0.2102	0.0417	2.2332	
Time to Convergence (Minutes)		**196**		**412**			

Table 4: Comparison between GMTM and SML, Two-Equation Model, Sample Size = 1000, Repetitions = 1000, Under the Assumption of Gamma (3, 2)

Parameters	**True Values**	**GMTM (MEAN)**	**GMTM (RMSE)**	**SML (MEAN)**	**SML (RMSE)**	**T-Test Values (RMSE)**	**% Censored Obs per Equation**
β_{11}	0.1000	0.0899	0.0190	0.0845	0.0237	2.3241	50.1%
β_{12}	0.2000	0.2070	0.0202	0.2167	0.0225	1.9846	
β_{21}	0.3000	0.2204	0.0179	0.2349	0.0398	2.0231	46.6%
β_{22}	0.4000	0.3928	0.0240	0.4166	0.0292	1.9975	
ρ	0.2000	0.1932	0.0068	0.1677	0.0323	2.4345	
Time to Convergence (Minutes)		**175**		**389**			

In terms of asymptotic efficiency of GMTM, more moment information is better than less, but this does not necessarily apply to finite sample properties of GMTM. There is a tendency for finite sample bias to eventually increase as an increasing number of estimating equations are used. To assess the statistical validity of the moment equations used in the GMTM approach,

Table 5: Comparison between GMTM and SML, Two-Equation Model, Sample Size = 1000, Repetitions = 1000, Under the Assumption of Gamma (4, 3)

Parameters	**True Values**	**GMTM (MEAN)**	**GMTM (RMSE)**	**SML (MEAN)**	**SML (RMSE)**	**T-Test Values (RMSE)**	**% Censored Obs per Equation**
β_{11}	0.1000	0.0394	0.0149	0.0404	0.0169	2.8940	49.4%
β_{12}	0.2000	0.2043	0.0155	0.2129	0.0162	2.7452	
β_{21}	0.3000	0.3199	0.0237	0.2747	0.0369	2.3180	46.3%
β_{22}	0.4000	0.3887	0.0105	0.4124	0.0396	2.8918	
ρ	0.2000	0.2061	0.0083	0.1917	0.0123	1.9986	
Time to Convergence (Minutes)		**191**		**394**			

we report a Chi-square test based on the asymptotic normal distribution of the estimating equations. In particular, defining $\mathbf{W}^* = \boldsymbol{Cov}\left(\mathbf{h}(\mathbf{Y},\mathbf{X},\boldsymbol{\theta})\right) = E\left(\mathbf{h}(\mathbf{Y},\mathbf{X},\boldsymbol{\theta})\mathbf{h}(\mathbf{Y},\mathbf{X},\boldsymbol{\theta})'\right)$, then $\mathbf{W}^{*-\frac{1}{2}}\mathbf{h}(\mathbf{Y},\mathbf{X},\boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0},\mathbf{I})$, and under the null hypothesis $H_o : E\left(\mathbf{h}(\mathbf{Y},\mathbf{X},\boldsymbol{\theta})\right) = \mathbf{0}$,

$$Q_{GMM(TEST)} = \mathbf{h}(\mathbf{Y},\mathbf{X},\hat{\boldsymbol{\theta}})'\mathbf{W}^{*-1}\mathbf{h}(\mathbf{Y},\mathbf{X},\boldsymbol{\theta}) \xrightarrow{d} \chi^2(m),$$

where *m* is the number estimating equations. Table 6 provides the GMTM moments validity tests for the two-equation model based on normal and Gamma data sampling distributions. The two-equation Tobit model has 44 degrees of freedom with a critical value of 60.48 at 95% significant level. The null hypothesis moment condition validity is not rejected for any of the data sampling scenarios, supporting the conclusion that that there is no statistical violation of the moment conditions used in GMTM estimation when using moments up to the third order.

Table 6: Chi-Square Moment Validity tests for the two-equation model with the assumption of different distributions on the error terms

Distributions	Sample size	Moment validity Chi-square test*
Normal (0,1)	100	15.2
Normal (0,1)	1000	13.3
Gamma(1,1)	1000	26.5
Gamma(3,2)	1000	20.4
Gamma(4,3)	1000	22.6

*Note: **The degrees of freedom is 44 and the critical value is 60.48 at the .05 significance level**

6.3 *Five-equation Tobit model*

Consider now the five-equation model presented in equation (6.1.2), with multivariate normally-distributed disturbances and a sample size of 1000 observations, the results of which are presented in Table 7. The t-test of the RMSEs shows that we reject the null hypothesis of equal RMSEs and conclude that GMTM provides an RMSE improvement over SML in this higher dimension model in spite of the large sample size.

The results may initially appear somewhat surprising since the GMTM uses less distributional information that SML, and SML is at least *asymptotically* efficient when the assumption of normality is correct. A partial explanation for these results, in addition to the ever-present discrepancies that can always exist between finite sample and asymptotic behavior, likely lies in the censoring rates for each equation in MVT systems coupled with the

Table 7: Comparison between GMTM and SML, Five-Equation Model, Sample Size = 1000, Repetitions = 1000, Under the Assumption of Normality

Parameter	True Value	GMTM (Mean)	GMTM (RMSE)	SML (RMSE)	SML (RMSE)	T-Test Values (RMSE)	% Censored Obs per Equation
β_{11}	0.1000	0.1022	0.0224	0.1086	0.0784	2.1354	48%
β_{12}	0.2000	0.2003	0.0130	0.2043	0.0253	2.0323	
β_{21}	0.3000	0.3032	0.0326	0.3134	0.0472	1.9981	53%
β_{22}	0.4000	0.4010	0.0142	0.4080	0.0568	2.3213	
β_{31}	0.5000	0.5021	0.0412	0.5141	0.0819	2.3421	42%
β_{32}	0.6000	0.6011	0.0241	0.6094	0.0964	2.4213	
β_{41}	0.7000	0.7021	0.0324	0.6982	0.0985	2.2243	51.5%
β_{42}	0.8000	0.8031	0.0143	0.8130	0.0432	2.5342	
β_{51}	0.9000	0.9100	0.0241	0.9181	0.0452	2.4325	47%
β_{52}	0.1100	0.1101	0.0128	0.1193	0.0845	2.3351	
ρ_{12}	0.1000	0.1014	0.0141	0.1207	0.0549	2.1352	
ρ_{13}	0.2000	0.2013	0.0136	0.2131	0.0562	2.1213	
ρ_{14}	0.3000	0.3009	0.0245	0.3090	0.0568	1.9875	
ρ_{15}	0.4000	0.4015	0.0314	0.3970	0.0454	2.0562	
ρ_{23}	0.6000	0.4022	0.0233	0.4103	0.0426	1.9895	
ρ_{24}	0.7000	0.7013	0.0138	0.7087	0.0542	2.1542	
ρ_{25}	0.2500	0.2500	0.0122	0.2530	0.0565	2.1438	
ρ_{34}	0.1100	0.1100	0.0130	0.1121	0.0423	2.0532	
ρ_{35}	0.1200	0.1190	0.0521	0.1300	0.0849	2.0462	
ρ_{45}	0.1500	0.1501	0.0233	0.1498	0.0989	2.1024	
Time to Convergence (Hours)		10		48			

substantially increased dimensionality and parameterization of the model. In the two-equation model censoring rates varied from 20% to 23%, and calculation of the probabilities associated with the discrete components of the likelihood function involve at most a two-dimensional integral. However, in the five-equation model censoring rates varied from 42% to 53% across the equations, and the discrete components require the solution of up to a five dimensional integration problem. As is well known, simulating a multidimensional integral, which represents the discontinuous likelihood component, becomes progressively more difficult and less accurate as dimensionality increases, and such inaccuracies can become more frequent as censoring become more prevalent.

As is the case with the two-equation simulation, GMTM exhibits advantages over SML in both tractability and speed of convergence in the five equation Tobit Model. In particular, given the higher dimensionality of the model, the speed advantage is substantially pronounced, where the GMTM convergence times are on average only 21% of SML convergence times.

7. CONCLUSIONS

This paper presents a Generalized Method of Truncated Moments (GMTM) method of estimating systems of censored equations that is computationally tractable, was shown to converge relatively quickly, is consistent, nearly asymptotically efficient, and is fully asymptotically efficient *relative* to the moment conditions used. The method uses univariate and bivariate moment conditions up to the third order to form the estimating equation information on which GMTM estimation is based, which is only a subset of the moments implied by a standard multivariate Tobit Model.

The GMTM estimator is compared to a conceptually asymptotically efficient full-information ML estimator obtained by simulating the likelihood (SML). Monte Carlo experiments reveal that under the assumption of normality and for small sample sizes, the GMTM estimator is actually more efficient than SML in our simulations, and is shown to be more robust to different skewed distributions. Moreover, in the presence of a high degree of censoring, a higher dimensional system, large sample size, and a multivariate normal data generating process, the GMTM estimator remains more efficient than SML.

A practical advantage of the GMTM approach over ML and SML is that in applications there is often insufficient information to accurately specify the parametric form of the likelihood function underlying the data generating process. The fact that not all of the moment behavior of the normal distribution is imposed in estimation was seen to afford the GMTM approach a robustness advantage when the distribution is misspecified. Given this situation, it could be worthwhile to consider expanding the implementation of the GMTM method to non-normally distributed data generating processes by defining the analogs of the truncated moment conditions used here, and potentially lending a degree of robustness as well as computational advantages relative to any baseline parametric distribution assumption for the data sampling process.

The results presented in this paper suggest that there are substantial potential gains to be had from pursuing a truncated moments GMM approach to the estimation of censored systems of equations. Future research on delineating methods for choosing the degree of moment reduction in the GMTM approach relative to tradeoffs in efficiency, robustness, tractability, and speed-to-convergence thus appear to be a worthwhile future endeavor.

A. APPENDIX: MOMENT DEFINITIONS

The complete moment results used in the implementation of the GMTM are presented in this appendix. Note that for the implementation of GMTM for all systems of equations of dimension three or greater, the pairwise results presented here can be used in forming all of the needed estimating equation information by simply interpreting $\left(y_1, y_2\right)$ as referring to $\left(y_i, y_j\right)$, for $i \neq j$, and $i \textit{ and } j \in \{1,2,...,J\}$.

Let $(y_1, y_2) \sim N\left(\begin{pmatrix} X_1\beta_1 \\ X_2\beta_2 \end{pmatrix}, \Sigma\right)$, and assume that the random variables are truncated as $y_1 \geq 0,\ y_2 \geq 0$. We can standardize the random variables so that we are dealing with correlated standard normal random variables as

$$
\begin{aligned}
y_1 \geq 0 \ &\Leftrightarrow\ z_1 = \frac{y_1 - x_1\beta_1}{\sigma_1} \geq \frac{-x_1\beta_1}{\sigma_1} = a_1 \\
y_2 \geq 0 \ &\Leftrightarrow\ z_2 = \frac{y_2 - x_2\beta_2}{\sigma_2} \geq \frac{-x_2\beta_2}{\sigma_2} = a_2
\end{aligned}
$$

Define the following for use in the expressions below:

$$
R = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},\ |R| = 1-\rho^2,\ R^{-1} = \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}\left(\frac{1}{1-\rho^2}\right) = \begin{pmatrix} (1-\rho^2)^{-1} & \dfrac{-\rho}{1-\rho^2} \\ \dfrac{-\rho}{1-\rho^2} & (1-\rho^2)^{-1} \end{pmatrix},
$$

$F^*(a_1, a_2) = \int_{a_1}^{\infty}\int_{a_2}^{\infty} BISN(z;\rho)dz$ represents the upper orthant integral of a bivariate standard normal random variable with correlation ρ, and ϕ and Φ represent the probability density function (PDF) and the cumulative distribution function (CDF) of a standard normal random variable, respectively.

The results provided in expressions A.1 – A.12 below are auxiliary results that are used in the definition of the main results provided in expressions A.13 – A.21. Additional details regarding the derivations underlying the expectation below are available from the authors.

A.1. *Auxiliary results*

$$A.1.\quad E(z_1 \mid z_1 \ge a_1, z_2 \ge a_2) = \int_{a_1}^{\infty} z_1 \int_{a_2}^{\infty} \frac{1}{(2\pi)(1-\rho^2)^{1/2}} e^{-1/2 z' R^{-1} z} dz_2 dz_1$$

$$A.2.\quad E(z_1 \mid z_1 \ge a_1, z_2 \ge a_2) = \int_{a_1}^{\infty} z_1 \int_{a_2}^{\infty} \frac{1}{(2\pi)(1-\rho^2)^{1/2}} \exp\left(\frac{-1}{2(1-\rho^2)}\left[z_1^2 + z_2^2 - 2\rho z_1 z_2\right]\right) dz_2 dz_1$$

$$= \frac{1}{F^*(a_1,a_2)}\left[\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \rho\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\right]$$

$$A.3.\quad E(z_1^2 \mid z_1 \ge a_1, z_2 \ge a_2) = 1 + \frac{1}{F^*(a_1,a_2)}\left[\begin{array}{l} a_1\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \rho^2 a_2\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) \\ +\rho\frac{(1-\rho^2)}{\sqrt{(1-\rho^2)}}\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) \end{array}\right]$$

$$A.4.\quad E(z_1^3 \mid z_1 \ge a_1, z_2 \ge a_2) = \frac{1}{F^*(a_1,a_2)}\left[\begin{array}{l} 3\times\left[\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \rho\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\right] \\ +\rho^3 a_2^2\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) + 2\rho^2 a_2\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\left(\frac{1-\rho^2}{\sqrt{1-\rho^2}}\right) \\ +\rho\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\left(1-\rho^2\right) + a_1^2\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) \end{array}\right]$$

$$A.5.\quad E(z_1z_2 \mid z_1 \geq a_1, z_2 \geq a_2) = \rho + \frac{1}{F^*(a_1,a_2)}\left[\begin{array}{l} \rho a_1\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \rho a_2\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) \\ +\left(\frac{1-\rho^2}{\sqrt{1-\rho^2}}\right)\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) \end{array}\right]$$

$$A.6.\quad E(z_1z_2^2 \mid z_1 \geq a_1, z_2 \geq a_2) = \frac{1}{F^*(a_1,a_2)}\left[\begin{array}{l} \left[\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \rho\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\right] \\ +2\rho\left[\rho\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\right] \\ +\rho^2a_1^2\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \rho a_1\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)\left(\frac{1-\rho^2}{\sqrt{1-\rho^2}}\right) \\ +\rho a_2^2\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) + a_2\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\left(\frac{1-\rho^2}{\sqrt{1-\rho^2}}\right) \end{array}\right]$$

$$A.7.\quad E(z_1^2z_2 \mid z_1 \geq a_1, z_2 \geq a_2) = \frac{1}{F^*(a_1,a_2)}\left[\begin{array}{l} 2\rho\left[\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \rho\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\right] \\ +\left[\rho\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\right] \\ +\rho^2a_2^2\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) + \rho a_2\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\left(\frac{1-\rho^2}{\sqrt{1-\rho^2}}\right) \\ +\rho a_1^2\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + a_1\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)\left(\frac{1-\rho^2}{\sqrt{1-\rho^2}}\right) \end{array}\right]$$

$$A.8.\quad E(z_2 \mid z_1 \geq a_1, z_2 \geq a_2) = \int_{a_2}^{\infty} z_2 \int_{a_1}^{\infty} \frac{1}{(2\pi)(1-\rho^2)^{1/2}} \exp\left(\frac{-1}{2(1-\rho^2)}\left[z_1^2 + z_2^2 - 2\rho z_1 z_2\right]\right) dz_1 dz_2$$

$$= \frac{1}{F^*(a_1,a_2)}\left[\rho\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\right]$$

$$A.9.\quad E(z_2^2 \mid z_1 \geq a_1, z_2 \geq a_2) = 1 + \frac{1}{F^*(a_1,a_2)}\left[\begin{array}{l} \rho^2a_1\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + a_2\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) \\ +\rho\frac{(1-\rho^2)}{\sqrt{(1-\rho^2)}}\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) \end{array}\right]$$

$$A.10.\quad E(z_2^3 \mid z_1 \geq a_1, z_2 \geq a_2) = \frac{1}{F^*(a_1,a_2)} \left[\begin{array}{c} 3 \times \left[\rho\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) \right] \\ +\rho^3 a_1^2 \phi(a_1)\Phi\left(\frac{a_2-\rho a_2}{(1-\rho^2)^{1/2}}\right) + 2\rho^2 a_1 \phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)\left(\frac{1-\rho^2}{\sqrt{1-\rho^2}}\right) \\ +\rho\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)\left(1-\rho^2\right) + a_2^2\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) \end{array} \right]$$

$$A.11.\quad E(z_2 z_1^2 \mid z_1 \geq a_1, z_2 \geq a_2) = \frac{1}{F^*(a_1,a_2)} \left[\begin{array}{c} 2\rho \left[\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \rho\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) \right] \\ + \left[\rho\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) \right] \\ +\rho^2 a_2^2 \phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) + \rho a_2 \phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\left(\frac{1-\rho^2}{\sqrt{1-\rho^2}}\right) \\ +\rho a_1^2 \phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + a_1 \phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)\left(\frac{1-\rho^2}{\sqrt{1-\rho^2}}\right) \end{array} \right]$$

$$A.12.\quad E(z_2^2 z_1 \mid z_1 \geq a_1, z_2 \geq a_2) = \frac{1}{F^*(a_1,a_2)} \left[\begin{array}{c} \left[\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \rho\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) \right] \\ +2\rho \left[\rho\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) \right] \\ +\rho^2 a_1^2 \phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \rho a_1 \phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)\left(\frac{1-\rho^2}{\sqrt{1-\rho^2}}\right) \\ +\rho a_2^2 \phi(a_2)\Phi\left(\frac{a_1-\rho a_1}{(1-\rho^2)^{1/2}}\right) + a_2 \phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\left(\frac{1-\rho^2}{\sqrt{1-\rho^2}}\right) \end{array} \right]$$

A.2. *Main results*

$$A.13.\quad E(y_1 \mid y_1 \geq 0, y_2 \geq 0) = \frac{\sigma_1}{F^*(a_1,a_2)} \left[\begin{array}{c} \phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \\ \rho\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) \end{array} \right] + x_1\beta_1$$

$$A.14.\quad E(y_2 \mid y_1 \ge 0, y_2 \ge 0) = \frac{\sigma_2}{F^*(a_1,a_2)}\begin{bmatrix} \rho\phi(a_1)\Phi\left(\dfrac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)+ \\ \phi(a_2)\Phi\left(\dfrac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\end{bmatrix} + x_2\beta_2$$

$$\begin{aligned} A.15.\quad E(y_1^2 \mid y_1 \ge 0, y_2 \ge 0) &= E((\sigma_1 z_1 + x_1\beta_1)^2 \mid z_1 \ge a_1, z_2 \ge a_2) \\ &= E((\sigma_1^2 z_1^2 + 2\left(x_1\beta_1\right)\sigma_1 z_1 + \left(x_1\beta_1\right)^2) \mid z_1 \ge a_1, z_2 \ge a_2) \\ &= \sigma_1^2 E(z_1^2 \mid z_1 \ge a_1, z_2 \ge a_2) + 2\sigma_1\left(x_1\beta_1\right)E(z_1 \mid z_1 \ge a_1, z_2 \ge a_2) + \left(x_1\beta_1\right)^2 \end{aligned}$$

$$\begin{aligned} A.16.\quad E(y_2^2 \mid y_1 \ge 0, y_2 \ge 0) &= E((\sigma_2 z_2 + x_2\beta_2)^2 \mid z_1 \ge a_1, z_2 \ge a_2) \\ &= E((\sigma_2^2 z_2^2 + 2\sigma_2(x_2\beta_2)z_2 + (x_2\beta_2)^2 \mid z_1 \ge a_1, z_2 \ge a_2) \\ &= \sigma_2^2 E(z_2^2 \mid z_1 \ge a_1, z_2 \ge a_2) + 2\sigma_2(x_2\beta_2)E(z_2 \mid z_1 \ge a_1, z_2 \ge a_2) + (x_2\beta_2)^2 \end{aligned}$$

$$\begin{aligned} A.17.\quad E(y_1^3 \mid y_1 \ge 0, y_2 \ge 0) &= E((\sigma_1 z_1 + x_1\beta_1)^3 \mid z_1 \ge a_1, z_2 \ge a_2) \\ &= E((\sigma_1^3 z_1^3 + 3\left(\sigma_1^2 z_1^2\right)\left(x_1\beta_1\right) + 3\left(\sigma_1 z_1\right)\left(x_1\beta_1\right)^2 + \left(x_1\beta_1\right)^3) \mid z_1 \ge a_1, z_2 \ge a_2) \\ &= \sigma_1^3 E(z_1^3 \mid z_1 \ge a_1, z_2 \ge a_2) + 3\sigma_1^2\left(x_1\beta_1\right)E(z_1^2 \mid z_1 \ge a_1, z_2 \ge a_2) \\ &\quad + 3\sigma_1\left(x_1\beta_1\right)^2 E(z_1 \mid z_1 \ge a_1, z_2 \ge a_2) + \left(x_1\beta_1\right)^3 \end{aligned}$$

$$\begin{aligned} A.18.\quad E(y_2^3 \mid y_1 \ge 0, y_2 \ge 0) &= E((\sigma_2 z_2 + x_2\beta_2)^3 \mid z_1 \ge a_1, z_2 \ge a_2) \\ &= E((\sigma_2^3 z_2^3 + 3\left(\sigma_2^2 z_2^2\right)\left(x_2\beta_2\right) + 3\left(\sigma_2 z_2\right)\left(x_2\beta_2\right)^2 + \left(x_2\beta_2\right)^3) \mid z_1 \ge a_1, z_2 \ge a_2) \\ &= \sigma_2^3 E(z_2^3 \mid z_1 \ge a_1, z_2 \ge a_2) + 3\sigma_2^2\left(x_2\beta_2\right)E(z_2^2 \mid z_1 \ge a_1, z_2 \ge a_2) \\ &\quad + 3\sigma_2\left(x_2\beta_2\right)^2 E(z_2 \mid z_1 \ge a_1, z_2 \ge a_2) + \left(x_2\beta_2\right)^3 \end{aligned}$$

$$\begin{aligned} A.19.\quad E(y_1 y_2 \mid y_1 \ge 0, y_2 \ge 0) &= E((\sigma_1 z_1 + \left(x_1\beta_1\right))(\sigma_2 z_2 + \left(x_2\beta_2\right)) \mid z_1 \ge a_1, z_2 \ge a_2) \\ &= E((\sigma_1\sigma_2 z_1 z_2 + \left(x_2\beta_2\right)\sigma_1 z_1 + \left(x_1\beta_1\right)\sigma_2 z_2 + \left(x_1\beta_1\right)\left(x_2\beta_2\right)) \mid z_1 \ge a_1, z_2 \ge a_2) \\ &= \sigma_1\sigma_2 E(z_1 z_2 \mid z_1 \ge a_1, z_2 \ge a_2) + \left(x_2\beta_2\right)\sigma_1 E(z_1 \mid z_1 \ge a_1, z_2 \ge a_2) \\ &\quad + \left(x_1\beta_1\right)\sigma_2 E(z_2 \mid z_1 \ge a_1, z_2 \ge a_2) + \left(x_1\beta_1\right)\left(x_2\beta_2\right) \end{aligned}$$

$$A.20.\quad E(y_1^2 y_2 \mid y_1 \geq 0, y_2 \geq 0) = E((\sigma_1 z_1 + (x_1\beta_1))^2(\sigma_2 z_2 + (x_2\beta_2)) \mid z_1 \geq a_1, z_2 \geq a_2)$$

$$= E(\sigma_1^2 z_1^2 \sigma_2 z_2 + 2\sigma_2 z_1 z_2 \sigma_1 (x_1\beta_1) + \sigma_2 z_2 (x_1\beta_1)^2 + \sigma_1^2 z_1^2 (x_2\beta_2)$$

$$+ 2\sigma_1 z_1 (x_1\beta_1)(x_2\beta_2) + (x_1\beta_1)^2 (x_2\beta_2)) \mid z_1 \geq a_1, z_2 \geq a_2)$$

$$= \sigma_1^2 \sigma_2 E(z_1^2 z_2 \mid z_1 \geq a_1, z_2 \geq a_2) + 2\sigma_1\sigma_2 (x_1\beta_1) E(z_1 z_2 \mid z_1 \geq a_1, z_2 \geq a_2)$$

$$+ \sigma_2 (x_1\beta_1)^2 E(z_2 \mid z_1 \geq a_1, z_2 \geq a_2) + \sigma_1^2 (x_2\beta_2) E(z_1^2 \mid z_1 \geq a_1, z_2 \geq a_2)$$

$$+ 2\sigma_1 (x_1\beta_1)(x_2\beta_2) E(z_1 \mid z_1 \geq a_1, z_2 \geq a_2) + (x_1\beta_1)^2 (x_2\beta_2)$$

$$A.21.\quad E(y_1 y_2^2 \mid y_1 \geq 0, y_2 \geq 0) = E((\sigma_1 z_1 + (x_1\beta_1))(\sigma_2 z_2 + (x_2\beta_2))^2 \mid z_1 \geq a_1, z_2 \geq a_2)$$

$$= E(\sigma_2^2 z_1 z_2^2 \sigma_1 + 2\sigma_1 z_1 \sigma_2 z_2 (x_2\beta_2) + \sigma_1 z_1 (x_2\beta_2)^2 + \sigma_2^2 z_2^2 (x_1\beta_1)$$

$$+ 2\sigma_2 z_2 (x_2\beta_2)(x_1\beta_1) + (x_2\beta_2)^2 (x_1\beta_1)) \mid z_1 \geq a_1, z_2 \geq a_2)$$

$$= \sigma_2^2 \sigma_1 E(z_1 z_2^2 \mid z_1 \geq a_1, z_2 \geq a_2) + 2\sigma_1\sigma_2 (x_2\beta_2) E(z_1 z_2 \mid z_1 \geq a_1, z_2 \geq a_2)$$

$$+ \sigma_1 (x_2\beta_2)^2 E(z_1 \mid z_1 \geq a_1, z_2 \geq a_2) + \sigma_2^2 (x_1\beta_1) E(z_2^2 \mid z_1 \geq a_1, z_2 \geq a_2)$$

$$+ 2\sigma_2 (x_1\beta_1)(x_2\beta_2) E(z_2 \mid z_1 \geq a_1, z_2 \geq a_2) + (x_2\beta_2)^2 (x_1\beta_1)$$

REFERENCES

Akhiezer, N. I. (1965) "The classical moment problem and some related questions in analysis", translated by N. Kemmer, Hafner Publishing Co., New York.

Amemiya, T., (1973) "Regression analysis when the dependent variable is truncated normal" Econometrica, 41, 997–1016

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., (1999) "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation," Econometrica, 67, 543–564

Berg, C., (1995) "Indeterminate moment problems and the theory of entire functions", J. Comput. Appl. Math. 65 (1995), no. 1, 27–55.

Blundell, R.W. and Meghir C., (1987) "Bivariate Alternatives to the Univariate Tobit Model," Journal of Econometrics, 34, 179–200

Chiang, J and Lee L., (1992) Discrete /continuous modes of consumer demand with binding non-negativity constrains," Journal of econometrics, 23, 59-80

Chung, C.E, and Goldberger, A.S., (1984) "Proportional Projections in Limited Dependent Variable Models," Econometrica , 52, 531–34

Chavas, J.P., and Perali, F., (2000) "Estimation of Censored Demand Equations from Large Cross-Sectional Data," Amer. J. Agr. Econ. 82, 1022–1037

Curto, R. E. and Fialkow, L. A., (1991) "Recursiveness, positivity, and truncated moment Problems", Houston J. Math. 17 , 603–635.

Curto, R. E. and Fialkow, L. A., (2005) "Truncated K-moment problems in several variables". Journal of Operator Theory, 54, pp. 189-226.

Deaton, A., (1986) "Demand Analysis, Handbook of Econometrics," vol. 3, chap. 30, pp. 1769–1839

Fialkow, L. A., (2008) "Truncated multivariable moment problems with finite variety", J. Operator Theory 60, 343–377.

Fuglede, B. (1983) "The multidimensional moment problem", Exposition. Math. **1**, no. 1, 47–65.

Genz, A. (1992), "Numerical Computation of Multivariate Normal Probabilities", Journal of Computational and Graphical Statistics, 1, pp. 141-149.

Greene, W.H., (1981) "On the Asymptotic Bias of the Ordinary Least Squares Estimator of the Tobit Model," Econometrica, 49, 505–13

Hajivassiliou, V. and Ruud, P., (1994) "Classical estimation methods for LDV models using simulation," In Handbook of Econometrics, Vol. 4, Chap. 40, North Holland, Amsterdam

Hajivassiliou, V., (1993b) "Simulation estimation methods for limited dependent variables," Handbook of statistics, vol.2, 519-543

Hajivassiliou, V.A., D. McFadden, and Ruud, P., (1996) "Simulation of Multivariate Normal Rectangle Probabilities and Their Derivatives: Theoretical and Computational Results," J. Econometrics, 72, 85–134

Hansen, L. and Singleton, (1982) "Generalized Instrumental Variable Estimators of Nonlinear Rational Expectations Models," Econometrica, 50, 1269–1286

Heckman, J. J., (1976*b*) "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and A Simple Estimator for Such Models," Annals Econ. and Soc. Meas., 5, 475–92

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., (1989) "Modeling Individual Choice," Oxford: UK Basil Blackwell

Shohat, J. A. and Tamarkin, J. D. , (1943) “The Problem of Moments”, American Mathematical Society Mathematical surveys, vol. II, American Mathematical Society, New York.

Stochel, J. (2001) “Solving the truncated moment problem solves the full moment problem”, Glasgow Mathematical Journal, 43, pp 335-341.

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