ELSEVIER

Journal of Econometrics 84 (1998) 383–400

JOURNAL OF
Econometrics

Estimation of censored linear errors-in-variables models

Liqun Wang

Institute for Statistics and Econometrics, University of Basel, Holbeinstrasse 12, CH-4051 Basel, Switzerland

Received 1 March 1993; revised 1 May 1997

Abstract

This paper deals with a linear errors-in-variables model where the dependent variable is censored. A two-step procedure is proposed to estimate the model and the corresponding asymptotic covariance matrices are derived. The framework covers the usual (error-free) Tobit model as a special case. It is shown that, under normality and a certain identifying condition, this model can be uniquely reduced to an error-free censored regression model and, hence, the existing estimators for the Tobit model can be used to obtain estimators for this model. In particular, the maximum-likelihood estimator is derived in this way. The small-sample behavior of the two estimators and their sensitivities to misspecified identifying information are studied through Monte-Carlo simulations. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Tobit model; Measurement errors; Limited dependent variable; Moment estimation; Maximum likelihood; Identification

JEL classification: C13; C24

1. Introduction

The censored regression models (Tobit models) are widely used in econometrics, biometrics and many other fields. See, e.g., Heckman and MaCurdy (1986), Killingsworth and Heckman (1986) and Pencavel (1986) for the applications of these models in labor econometrics. The statistical theories and methods for these models can be found in Amemiya (1984, 1985), Maddala (1985) and Greene (1993). Usually, it is assumed in these models that the explanatory variables are

The author wishes to thank Shoaleh Ghashghaie, Klaus Poetzelberger, Wolfgang Polasek, James Chu, Cheng Hsiao, Andrew Weiss, Quang Wuong, Richard Blundell and a referee for helpful discussions and suggestions. Financial support from Swiss National Science Foundation is gratefully acknowledged.

PII S0304-4076(97)00093-6

bounded constants and are exactly observed. Obviously, such assumptions are not always appropriate in many real problems and may result in inaccurate and inconsistent estimates. Recently, several authors have considered the censored regression models or other limited dependent variable models with measurement errors. Hsiao (1991) studied a class of binary choice models where the explanatory variables are measured with errors. Weiss (1993) investigated the least absolute deviation estimators of a censored linear errors-in-variables model when certain instrumental variables are available. Whereas Colombi (1993) treated a more general class of latent variables models. For the censored regression models with errors-in-variables, however, statistical theories of the most commonly used moment estimator and maximum-likelihood estimator are not yet available. The objective of the present paper is to fill in this gap.

Specifically, we consider the following censored linear errors-in-variables model:

$$\eta_t = \beta_1 + \beta_2'\xi_t + u_t, \quad y_t = \max\{\eta_t, 0\}, \quad x_t = \xi_t + v_t, \tag{1}$$

where $\eta_t \in \mathbb{R}$, $\xi_t \in \mathbb{R}^k$ are the unobserved variables, β_1, β_2 the regression parameters, y_t, x_t the observed variables and u_t, v_t the errors (here we use the word *unobserved* also for partly unobserved variable). Furthermore, we assume that u_t, v_t and ξ_t are independently and normally distributed with means 0, 0, μ_ξ and variances σ_u, Σ_v, Σ_ξ, respectively.

The major difference between model (1) and the ordinary (error-free) Tobit model is that some or all components of the explanatory variable ξ_t are subject to measurement errors. As is well-known, this feature causes many difficulties and complexities in conducting the statistical analysis of the model, because now the x_t's are no longer constants and, as a result, the distributions of the x_t's enter the likelihood function of the model. In this paper we show that, given the normality assumption and certain identifying information, model (1) can be uniquely reduced to an error-free censored regression model and, consequently, the estimators of this model can be derived through the existing estimators of the Tobit model. In particular, the maximum-likelihood estimator (MLE) is derived using this approach. We also propose an alternative approach to derive the two-step moment estimators (TME) and their asymptotic covariance matrices. It is worth noting that, although the method of moments is widely applied in many estimation procedures, the asymptotic covariance matrix of the moment estimator for the usual Tobit model is not yet available. To the best of author's knowledge, the only recent work dealing with this aspect in the econometric literature is Greene (1983), where only the asymptotic covariance matrix of the moment estimator of the slope parameter β_2 is given. As a by-product, therefore, the asymptotic results in Section 3 provide a complete formula for the asymptotic covariance matrix of the moment estimator of all parameters in the Tobit model, which is considered as a special case of model (1) where the measurement errors have zero covariance matrix $\Sigma_v = 0$. As for the a priori identifying condition, instead of assuming the existence of certain instrumental variables we assume that the noise-to-signal ratio $\Delta = \Sigma_\xi^{-1}\Sigma_v$ is known.

The paper is organized as follows. Section 2 is concerned with the identifiability of model (1). Section 3 shows how model (1) can be uniquely reduced to an error-free model and how the estimators are derived in general. In Section 4 two-step moment estimators are proposed and their asymptotic covariance matrices are derived. The maximum-likelihood estimator is derived in Section 5. The small sample behavior of these two procedures and their sensitivities to the misspecified identifying information Δ are investigated through Monte-Carlo simulations in Section 6. The conclusions are in Section 7.

2. Identifiability

It is well-known that the usual linear normal errors-in-variables model is not identifiable and hence the model cannot be estimated consistently. In this section, we examine the identifiability of model (1).

For identifiability we adopt the definition of Hsiao (1983) or Fuller (1987), Section 1.1.3. Formally, denote the sample $z = (y_t, x_t', \ t = 1, 2, \ldots, T)$, where T is the sample size, and suppose the sample distribution function $F(z \mid \theta)$ is known up to an unknown parameter $\theta \in \mathbb{R}^p$. Let $\Theta \subseteq \mathbb{R}^p$ denote the parameter space. Then the model is said to be identifiable, if for any $\theta_1, \theta_2 \in \Theta$, $F(z \mid \theta_1) \equiv F(z \mid \theta_2)$ implies $\theta_1 = \theta_2$. A parameter in the model (a component of $\theta \in \Theta$) is said to be identified, if it is uniquely determined by the sample distribution.

Now in model (1) the sample distribution is a product of the distributions of (y_t, x_t') because the sample is i.i.d. From (1) the distribution of (y_t, x_t') is completely determined by the distribution of (η_t, x_t'), which is in turn completely determined by the first and second moments of (η_t, x_t') under normality. Further, the first and second moments of (η_t, x_t') are related to the model parameters through the equations

$$\begin{aligned} &\mu_\eta = \beta_1 + \beta_2'\mu_\xi, \quad \mu_x = \mu_\xi, \quad \sigma_\eta = \beta_2'\Sigma_\xi\beta_2 + \sigma_u, \\ &\sigma_{x\eta} = \Sigma_\xi\beta_2, \quad \Sigma_x = \Sigma_\xi + \Sigma_v. \end{aligned} \tag{2}$$

From (2) it is clear that there are more than one sets of values of parameters $(\beta_1, \beta_2, \sigma_u, \Sigma_\xi, \Sigma_v)$ which are compatible with the same set of values of the parameters on the left-hand side of (2) and, therefore, are compatible with the same sample distribution of (y_t, x_t'). As a result, in general, model (1) is not identifiable.

Next we show that, except $(\beta_1, \beta_2, \sigma_u, \Sigma_\xi, \Sigma_v)$, all other parameters in the model are identified. Indeed, the identifiability of μ_ξ is clear. To see that μ_η, σ_η and $\sigma_{x\eta}$

are also identified, let $\delta = \mu_\eta / \sqrt{\sigma_\eta}$. Then, by straightforward integration we have

$$\begin{aligned} E(y_t) &= \Phi(\delta)E(y_t \mid y_t > 0), \\ E(y_t \mid y_t > 0) &= \mu_\eta + \sqrt{\sigma_\eta}\phi(\delta)/\Phi(\delta), \\ E(x_t y_t \mid y_t > 0) &= \sigma_{x\eta} + \mu_x E(y_t \mid y_t > 0), \end{aligned} \tag{3}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal distribution and density functions. It is apparent from (3) that μ_η, σ_η and $\sigma_{x\eta}$ are uniquely determined by the moments of (y_t, x_t') which appear in (3).

It follows that the sample distribution of (y_t, x_t') and the first two moments of (η_t, x_t') are mutually uniquely determined. As a result, the identifiability of model (1) is equivalent to the unique determination of the model parameters by Eqs. (2), given its left-hand side. A simple counting process shows that (2) contains $(k+1)(k+4)/2$ independent equations but $(k+1)(k+2)$ free parameters on the right-hand side. Thus, as in the case of usual linear errors-in-variables models, a priori restrictions on the model parameters $(\beta_1, \beta_2, \sigma_u, \Sigma_\xi, \Sigma_v)$ are needed to ensure identifiability.

In practical applications the a priori identifying information is usually provided in terms of at least $k(k+1)/2$ linear restrictions on the parameters, e.g., that the variance ratio $\sigma_u^{-1}\Sigma_v$ or the so-called reliability ratio $\kappa = \Sigma_x^{-1}\Sigma_\xi$ is known or may be determined previously. The latter is equivalent to the condition that the noise-to-signal ratio $\Lambda = \Sigma_\xi^{-1}\Sigma_v$ is known because $\kappa = (I + \Lambda)^{-1}$. This information can be obtained in many situations when, e.g., validation data, panel data or repeated sampling are available. For more discussions see, e.g., Jaech (1985), Fuller (1987) and Gleser (1992). In the rest of this paper we assume that $\Lambda = \Sigma_\xi^{-1}\Sigma_v$ is known.

3. Model reduction and estimation

In this section we show that model (1) can be reduced to an error-free form. Indeed, let $A = \Sigma_\xi \Sigma_x^{-1}$ and $b = (I - A)\mu_x$, then

$$\xi_t = Ax_t + b + e_t, \tag{4}$$

where $e_t = (I - A)(x_t - \mu_x) - v_t$ is independent of x_t under normality. Substituting (4) into the first equation of (1) and combining it with the second equation of (1) result in the familiar form of a censored regression model:

$$\begin{aligned} \eta_t &= \gamma_1 + \gamma_2' x_t + w_t, \\ y_t &= \max\{\eta_t, 0\}, \end{aligned} \tag{5}$$

where $w_t = u_t + \beta_2' e_t$ has distribution $N(0, \sigma_w)$ and is independent of x_t. The relations between the new parameters $(\gamma_1, \gamma_2, \sigma_w, \mu_x, \Sigma_x)$ and the original ones

$(\beta_1, \beta_2, \sigma_u, \mu_\xi, \Sigma_\xi, \Sigma_v)$ are given by

$$\begin{aligned} &\beta_1 = \gamma_1 - \mu_x' \Delta \gamma_2, \quad \beta_2 = (I + \Delta)\gamma_2, \quad \sigma_u = \sigma_w - \gamma_2' \Sigma_x \Delta \gamma_2, \\ &\mu_\xi = \mu_x, \quad \Sigma_\xi = \Sigma_x (I + \Delta)^{-1}, \end{aligned} \tag{6}$$

where $\Delta = \Sigma_\xi^{-1} \Sigma_v$. Clearly, the mapping (6) is one-to-one. Consequently, any estimator of model (5) implies a corresponding estimator of model (1). Using this approach it is also possible to derive the asymptotic bias of the estimator of model (1) when the identifying information Δ is misspecified. For instance, if $\hat{\psi} = (\hat{\gamma}_1, \hat{\gamma}_2', \hat{\sigma}_w)'$ is a consistent estimator of model (5) and $\tilde{\theta} = (\tilde{\beta}_1, \tilde{\beta}_2', \tilde{\sigma}_u)'$ is obtained via (6) where, instead of Δ, a wrong $\tilde{\Delta}$ is used. Then, the asymptotic bias of $\tilde{\theta}$ is given by

$$\begin{aligned} \operatorname{plim} \tilde{\beta}_1 &= \beta_1 + \mu_x' (\Delta - \tilde{\Delta})(I + \Delta)^{-1} \beta_2, \\ \operatorname{plim} \tilde{\beta}_2 &= \beta_2 - (\Delta - \tilde{\Delta})(I + \Delta)^{-1} \beta_2, \\ \operatorname{plim} \tilde{\sigma}_u &= \sigma_u + \beta_2' \Sigma_\xi (\Delta - \tilde{\Delta})(I + \Delta)^{-1} \beta_2. \end{aligned} \tag{7}$$

From (7) we see that the estimation biases are of the same order as $\Delta - \tilde{\Delta}$ and, hence, can be significant if the amount of misspecification $\Delta - \tilde{\Delta}$ is not very small relative to $I + \Delta$. Furthermore, the slope parameter β_2 tends to be underestimated by underspecified Δ and overestimated by overspecified Δ, whereas the converse is true for β_1 and σ_u.

Finally, we note that model (5) is different from the ordinary Tobit model in that the x_t in (5) is a random variable and is unbounded under normality, whereas in the Tobit model it is assumed to be bounded constants. This fact should be taken into account in deriving the asymptotic covariance matrices of the estimators. In Section 5 we show how the maximum-likelihood estimator for model (1) can be derived in the way described in this section.

4. Two-step moment estimators (TME)

In this section, we consider the moment estimator of model (1). As has been discussed in Section 3, one way of deriving the moment estimator is to use (6) and the corresponding moment estimator of model (5). A technical difficulty with this approach, however, is that, in order to derive the asymptotic covariance matrix of the estimator, the corresponding asymptotic covariance matrix of the moment estimator of model (5) is needed, which is not yet available. To overcome this difficulty we propose a two-step procedure based on the discussion in Section 2: first, the first and second moments of (η_t, x_t') are estimated using (3); and then, the other parameters are estimated by solving Eqs. (2) with the left-hand side substituted through the sample moments and the estimates obtained from the first step.

Suppose the data (y_t, x_t'), $t = 1, 2, \ldots, T$, are given, in which T_0 y_t's are zero and $T_1 = T - T_0$ y_t's are positive. To avoid the trivial case we assume $0 < T_0 < T$. Further, denote the sample moments $\hat{\mu}_y = (1/T)\sum_{t=1}^T y_t$, $\hat{\mu}_x = (1/T)\sum_{t=1}^T x_t$ and $\hat{\Sigma}_x = (1/T)\sum_{t=1}^T (x_t - \bar{x})(x_t - \bar{x})'$. The conditional moments in (3) are denoted as μ_{y+}, μ_{y^2+} and μ_{xy+}, which are consistently estimated by the corresponding sample moments using the positive y_t's and the corresponding x_t's. These estimators are denoted analogously as $\hat{\mu}_{y+}$, $\hat{\mu}_{y^2+}$ and $\hat{\mu}_{xy+}$. Then, by definition, $\hat{\delta} = \Phi^{-1}(\hat{\mu}_y/\hat{\mu}_{y+}) = \Phi^{-1}(T_1/T)$ and from (3) we have

$$\hat{\mu}_\eta = \hat{\delta}\hat{\mu}_{y+}/(\hat{\delta} + \phi(\hat{\delta})/\Phi(\hat{\delta})), \quad \hat{\sigma}_\eta = (\hat{\mu}_\eta/\hat{\delta})^2, \quad \hat{\sigma}_{x\eta} = \hat{\mu}_{xy+} - \hat{\mu}_x\hat{\mu}_{y+}. \tag{8}$$

Substituting $\hat{\mu}_x$, $\hat{\Sigma}_x$ and (8) into the left-hand side of (2) and solving these equations we obtain

$$\begin{aligned} &\hat{\beta}_2 = \hat{\Sigma}_\xi^{-1}\hat{\sigma}_{x\eta}, \quad \hat{\beta}_1 = \hat{\mu}_\eta - \hat{\beta}_2'\hat{\mu}_\xi, \\ &\hat{\sigma}_u = \hat{\sigma}_\eta - \hat{\beta}_2'\hat{\sigma}_{x\eta}, \quad \hat{\mu}_\xi = \hat{\mu}_x, \quad \hat{\Sigma}_\xi = \hat{\Sigma}_x(I + \Delta)^{-1}. \end{aligned} \tag{9}$$

Clearly, all estimators given by (8) and (9) are strongly consistent because they are continuous functions of the sample moments. To derive the asymptotic normalities of these two-step moment estimators, we denote $\psi = (\mu_\eta, \sigma_\eta, \sigma_{x\eta}', \mu_x', \sigma_x')$ and $\theta = (\beta_1, \beta_2', \sigma_u, \mu_\xi', \sigma_\xi')'$, where

$$\sigma_x = \operatorname{vec}\Sigma_x = \mu_{x\otimes x} - \mu_x \otimes \mu_x,$$

$\sigma_\xi = \operatorname{vec}\Sigma_\xi$, vec is the usual column vectorization operator and $\otimes$ is the Kronecker product. Correspondingly, we denote the first-step estimators by $\hat{\psi}_{\mathrm{TM}} = (\hat{\mu}_\eta, \hat{\sigma}_\eta, \hat{\sigma}_{x\eta}', \hat{\mu}_x', \hat{\sigma}_x')'$ and the second-step estimators by $\hat{\theta}_{\mathrm{TM}} = (\hat{\beta}_1, \hat{\beta}_2', \hat{\sigma}_u, \hat{\mu}_\xi', \hat{\sigma}_\xi')'$. Then, as is shown in Appendix A, we have the following expressions:

$$\hat{\psi}_{\mathrm{TM}} - \psi_0 = \hat{A}\sum_{t=1}^T z_t/T, \tag{10}$$

and

$$\hat{\theta}_{\mathrm{TM}} - \theta_0 = \hat{B}(\hat{\psi}_{\mathrm{TM}} - \psi_0), \tag{11}$$

where ψ_0 and θ_0 correspond to the true parameters of model (1), $\hat{A}$ and $\hat{B}$ converge in probability to

$$A=\begin{pmatrix} \dfrac{\mu_{y^2+}}{\mu_y} & \dfrac{\mu_\eta}{\mu_{y+}\Phi(\delta)} & 0 & 0 & 0 \\ 2\sigma_\eta\left(\dfrac{\mu_{y^2+}}{\mu_\eta\mu_y}-\dfrac{1}{\delta\phi(\delta)}\right) & \dfrac{2\sigma_\eta}{\mu_{y+}\Phi(\delta)} & 0 & 0 & 0 \\ 0 & -\dfrac{\mu_x}{\Phi(\delta)} & \dfrac{1}{\Phi(\delta)}I_k & -\mu_{y+}I_k & 0 \\ 0 & 0 & 0 & I_k & 0 \\ 0 & 0 & 0 & -I_k\otimes\mu_x-\mu_x\otimes I_k & I_k \end{pmatrix} \tag{12}$$

and

$$B=\begin{pmatrix} 1 & 0 & -\mu_x'\Sigma_\xi^{-1} & -\beta_2' & \sigma_{x\eta}'\Sigma_x^{-1}\otimes\mu_x'\Sigma_\xi^{-1} \\ 0 & 0 & \Sigma_\xi^{-1} & 0 & -\sigma_{x\eta}'\Sigma_x^{-1}\otimes\Sigma_\xi^{-1} \\ 0 & 1 & -2\beta_2' & 0 & \sigma_{x\eta}'\Sigma_x^{-1}\otimes\beta_2' \\ 0 & 0 & 0 & I_k & 0 \\ 0 & 0 & 0 & 0 & \Sigma_\xi\Sigma_x^{-1}\otimes I_k \end{pmatrix}, \tag{13}$$

respectively,

$$z_t=(s_t-\Phi(\delta),\, s_t(y_t-\mu_{y+}),\, s_t(x_ty_t-\mu_{xy+})',\, (x_t-\mu_x)',(x_t\otimes x_t-\mu_{x\otimes x})')'$$

and

$$s_t=\begin{cases}1, & y_t>0,\\ 0, & y_t=0.\end{cases}$$

Since z_t are i.i.d., by a central limit theorem,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}z_t \xrightarrow{L} \mathrm{N}(0,\Sigma_z) \tag{14}$$

with

$$\Sigma_z=Ez_tz_t'=\Phi(\delta)(\Sigma_1,\Sigma_2),$$

where

$$\Sigma_1 = \begin{pmatrix} 1-\Phi(\delta) & * & * \\ 0 & \sigma_{y+} & * \\ 0 & \mu_{xy^2+} - \mu_{y+}\mu_{xy+} & \Sigma_{xy+} \\ \mu_{x+} - \mu_x & \mu_{xy+} - \mu_{x+}\mu_{y+} & \mu_{xx'y+} - \mu_{x+}\mu'_{xy+} \\ \mu_{x\otimes x+} - \mu_{x\otimes x} & \mu_{(x\otimes x)y+} - \mu_{x\otimes x+}\mu_{y+} & \mu_{(x\otimes x)x'y+} - \mu_{x\otimes x+}\mu'_{xy+} \end{pmatrix}$$

and

$$\Sigma_2 = \begin{pmatrix} * & * \\ * & * \\ * & * \\ \Sigma_x/\Phi(\delta) & * \\ (\mu_{(x\otimes x)x'} - \mu_{x\otimes x}\mu'_x)/\Phi(\delta) & \Sigma_{x\otimes x}/\Phi(\delta) \end{pmatrix}.$$

The following results follow then from (10), (11), (14) and Slutsky Theorem (Amemiya, 1985).

Theorem 1. Suppose $\Sigma_\xi > 0$ *and* $\Lambda = \Sigma_\xi^{-1}\Sigma_v$ *is known. Then,*
(1) *all estimators given by* (8)–(9) *are strongly consistent;*
(2) $\sqrt{T}(\hat{\psi}_{TM} - \psi_0) \xrightarrow{L} N(0, A\Sigma_z A')$, *where* A *is given by* (12);
(3) $\sqrt{T}(\hat{\theta}_{TM} - \theta_0) \xrightarrow{L} N(0, BA\Sigma_z A'B')$, *where* B *is given by* (13).

Remark 1. The two-step procedure used in this section may be similarly applied to the case where instead of Λ the variance ratio $\sigma_u^{-1}\Sigma_v$ is known. The only difference is that the second-step estimators should be calculated similarly as in Fuller (1987), Section 1.3. The asymptotic results of the estimators may be established analogously to Theorem 1. Such results for a simple model with $k = 1$ are given by Theorem 1.3.1 of Fuller (1987).

Remark 2. If $\Lambda = 0$, then $\Sigma_\xi = \Sigma_x$ and model (1) reduces to the usual error-free Tobit model. Thus, the asymptotic covariance matrices given in Theorem 1 apply to the moment estimators of the Tobit model as well. These formulae have not been given in the literature, though the moment estimators are widely used in estimation procedures.

5. Maximum-likelihood estimator (MLE)

In this section we derive the MLE of model (1) via the MLE of model (5). In particular, we show that the results of Olsen (1978) concerning the existence of the unique, global MLE for model (5) and the results of Amemiya (1973) concerning the asymptotic normality of the MLE can be used to obtain the corresponding results of the MLE for model (1).

Let the data be given as in Section 4. Without loss of generality, we assume that the first T_0 y_t's are zero and the last $T_1 = T - T_0$ y_t's are positive. Then the log-likelihood function of model (1) is, up to a constant,

$$L = \sum_{t=1}^{T_0} \log \Phi\left(-\frac{\gamma_1 + \gamma_2' x_t}{\sqrt{\sigma_w}}\right) - \frac{T_1}{2}\log \sigma_w - \frac{1}{2\sigma_w}\sum_{t=T_0+1}^{T}(y_t - \gamma_1 - \gamma_2' x_t)^2$$

$$-\frac{T}{2}\log\det \Sigma_x - \frac{1}{2}\sum_{t=1}^{T}(x_t - \mu_x)'\Sigma_x^{-1}(x_t - \mu_x), \tag{15}$$

where $(\gamma_1, \gamma_2, \sigma_w, \mu_x, \Sigma_x)$ are given by (6). It is clear that the first part (the first line) of (15) is just the conditional log-likelihood function of model (5) and does not involve μ_x and Σ_x. As a result, the MLE of μ_x and Σ_x are given by the corresponding sample moments and the MLE of μ_ξ and Σ_ξ are, therefore, identical with the TME in (9). The MLE of $(\beta_1, \beta_2, \sigma_u)$ may be obtained via (6) and the MLE of $(\gamma_1, \gamma_2, \sigma_w)$ for model (5). Using the reparameterization $\tau = 1/\sqrt{\sigma_w}$ and $\alpha = (\tau\gamma_1, \tau\gamma_2')'$, Olsen (1978) shows that the conditional log-likelihood function (the first line of (15))

$$L_c(\psi) = \sum_{t=1}^{T_0} \log \Phi(-\alpha' \tilde{x}_t) + T_1 \log \tau - \frac{1}{2}\psi' Z' Z \psi \tag{16}$$

is globally concave in $\psi = (\alpha', \tau)' \in \Psi = \mathbb{R}^{k+1} \times \mathbb{R}_+$, where $\mathbb{R}_+ = (0, +\infty)$, $\tilde{x}_t = (1, x_t')'$, $Z = (X_1, -Y_1)$, $X_1 = (\tilde{x}_{T_0+1}, \tilde{x}_{T_0+2}, \ldots, \tilde{x}_T)'$ and $Y_1 = (y_{T_0+1}, y_{T_0+2}, \ldots, y_T)'$. The asymptotic normality of the MLE for the Tobit model is established by Amemiya (1973). Adapted to our situation where the x_t is a random variable, the asymptotic covariance matrix of $\hat{\psi}_{ML}$ obtained by maximizing $L_c(\psi)$ in (16) can be written as the inverse of

$$-\operatorname*{plim}_{T\to\infty} \frac{1}{T}\frac{\partial^2 L_c(\psi)}{\partial\psi\,\partial\psi'} = \Omega = \Omega_0 + \Phi(\delta)\Omega_1,$$

where $\delta = \mu_\eta / \sqrt{\sigma_\eta}$,

$$\Omega_0 = \begin{pmatrix} \Phi(-\delta)\mathrm{E}[\lambda_t(\lambda_t - \alpha'\tilde{x}_t)\tilde{x}_t\tilde{x}_t' \mid \eta_t \leqslant 0] & 0 \\ 0 & \Phi(\delta)/\tau^2 \end{pmatrix},$$

$$\Omega_1 = \begin{pmatrix} \mathrm{E}(\tilde{x}_t\tilde{x}_t' \mid y_t > 0) & -\mathrm{E}(\tilde{x}_t y_t \mid y_t > 0) \\ -\mathrm{E}(y_t\tilde{x}_t' \mid y_t > 0) & \mathrm{E}(y_t^2 \mid y_t > 0) \end{pmatrix},$$

and $\lambda_t = \phi(\alpha'\tilde{x}_t)/\Phi(-\alpha'\tilde{x}_t)$. The derivation is straightforward by using the standard results of normal distribution. An alternative and direct derivation of the asymptotic covariance matrix Ω is given by Wang (1994). Now, the MLE for $\theta = (\beta_1, \beta_2', \sigma_u)'$ is calculated according to

$$\begin{aligned} \beta_1 &= (\alpha_1 - \mu_x' \Delta \alpha_2)/\tau, \\ \beta_2 &= (I + \Delta)\alpha_2/\tau, \\ \sigma_u &= (1 - \alpha_2' \Sigma_x \Delta \alpha_2)/\tau^2. \end{aligned} \tag{17}$$

Let $\theta(\psi): \Psi \mapsto \Theta$ denote the mapping (17) and $\hat{\theta}_{\mathrm{ML}} = \theta(\hat{\psi}_{\mathrm{ML}})$. Then the consistency of $\hat{\theta}_{\mathrm{ML}}$ follows immediately from the continuity of $\theta(\psi)$. To show the asymptotic normality of $\hat{\theta}_{\mathrm{ML}}$, note that $\theta(\psi)$ is continuously differentiable and, hence, we have the first-order Taylor expansion

$$\hat{\theta}_{\mathrm{ML}} - \theta_0 = \frac{\partial \theta(\tilde{\psi})}{\partial \psi'}(\hat{\psi}_{\mathrm{ML}} - \psi_0),$$

where $\tilde{\psi}$ lies between $\hat{\psi}_{\mathrm{ML}}$ and ψ_0. The derivative $\partial\theta/\partial\psi'$ is obviously a continuous function of ψ and, therefore, converges in probability to the matrix

$$C = \sqrt{\sigma_u} \begin{pmatrix} 1 & -\mu_x'\Delta & -\beta_1 \\ 0 & I + \Delta & -\beta_2 \\ 0 & -2\beta_2'\Sigma_x & -2\sigma_u \end{pmatrix}. \tag{18}$$

Thus, we have the following results.

Theorem 2. Suppose $\sigma_u > 0$, $\Sigma_\xi > 0$ and $\Delta = \Sigma_\xi^{-1}\Sigma_v$ is given. Then,
(1) *the log-likelihood function L in (15) has a unique, finite, global maximum in the parameter space $\tilde{\Theta} = \mathbb{R}^{2k+1+k(k-1)/2} \times \mathbb{R}_+^{k+1}$;*
(2) *$\hat{\theta}_{\mathrm{ML}} \xrightarrow{\mathrm{P}} \theta_0$, where θ_0 is the true parameter of model (1);*
(3) *$\sqrt{T}(\hat{\theta}_{\mathrm{ML}} - \theta_0) \xrightarrow{\mathrm{L}} \mathrm{N}(0, C\Omega^{-1}C')$, where C is given by (18).*

Remark 3. The maximization of $L_c(\psi)$ may be carried out through standard numerical methods such as Newton–Raphson. The numerical calculation is straightforward as the first and second derivatives of $L_c(\psi)$ are available:

$$\frac{\partial L_c(\psi)}{\partial \psi} = \begin{pmatrix} -X_0'\lambda_0 \\ T_1/\tau \end{pmatrix} - Z'Z\psi$$

and

$$\frac{\partial^2 L_c(\psi)}{\partial \psi \partial \psi'} = -\begin{pmatrix} X_0'\Lambda_0 X_0 & 0 \\ 0 & T_1/\tau^2 \end{pmatrix} - Z'Z,$$

where $X_0 = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_{T_0})'$, $\lambda_0 = (\lambda_t,\ t = 1,2,\ldots,T_0)'$, $\lambda_t = \phi(\alpha'\tilde{x}_t)/\Phi(-\alpha'\tilde{x}_t)$ and Λ_0 is the diagonal matrix with the diagonal elements $\lambda_t(\lambda_t - \alpha'\tilde{x}_t)$, $t = 1,2,\ldots,T_0$. Since $L_c(\psi)$ is globally concave, the iteration may start at any finite point. However, a good starting point is important for rapid convergence. The TME of Section 4 may serve as initial values for the iterations. As is shown by the Monte-Carlo simulation study in the next section, for a simple model with $k = 1$ the MLE procedures using the Newton–Raphson algorithm and the TME as starting values may achieve rather satisfactory convergence after four or five iterations. Furthermore, the estimators $\hat{\psi}_1$ and $\hat{\theta}_1$ obtained after one iteration in the Newton–Raphson procedure have the same asymptotic distributions as the MLE $\hat{\psi}_{ML}$ and $\hat{\theta}_{ML}$, respectively.

6. Monte-Carlo simulations

Both the TME and the MLE of model (1) derived in previous sections are consistent and asymptotically normal under general conditions. In this section we study through Monte-Carlo simulations the behavior of the two procedures when the sample size is small or the a priori information Λ is misspecified. We consider a simple model with $k = 1$. In this case we use lower-case letters to denote all moments. The true values for the model are $\beta_1 = -6$, $\beta_2 = 0.6$, $\sigma_u = \sigma_v = 18$, $\mu_\xi = 20$ and $\sigma_\xi = 180$. Thus, the true noise-to-signal ratio is $\Lambda = \sigma_v/\sigma_\xi = 0.1$. We simulate the means and the mean absolute deviations

$$\mathrm{MAD}(\hat{\theta}_i) = \mathrm{E}|\hat{\theta}_i - \theta_i|, \quad i = 1,2,3$$

for the estimators $\hat{\theta} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}_u)'$. The TME are calculated according to (8)–(9), whereas the MLE are calculated by (17) and by maximizing the function L_c in (16) through the Newton-Raphson method. In each simulation $N = 1000$ replications have been carried out. The average amount of censored observations in the samples is about 25% and the average number of iterations in calculating the MLE is $I = 4$ with convergence criterion $\varepsilon = 10^{-5}$.

Table 1
TME and MLE for $\Delta = 0.1$ and various sample sizes T

T	20	30	40	50	100	200	400
$\beta_1 = -6$							
TME	−5.7304	−6.0763	−6.0413	−5.9857	−5.9937	−6.0078	−5.9979
MAD	2.5710	2.1882	1.8461	1.6334	1.3374	0.9007	0.5833
MLE	−6.1719	−6.2827	−6.1753	−6.0898	−6.0097	−6.0324	−6.0045
MAD	2.2758	1.8426	1.5501	1.3593	0.9878	0.6881	0.4918
$\beta_2 = 0.6$							
TME	0.5799	0.5156	0.5969	0.5942	0.5963	0.5992	0.5996
MAD	0.1026	0.0850	0.0722	0.0623	0.0515	0.0343	0.0227
MLE	0.6075	0.6110	0.6069	0.6025	0.5986	0.6013	0.6002
MAD	0.0930	0.0728	0.0616	0.0532	0.0387	0.0272	0.0192
$\sigma_u = 18$							
TME	20.8168	20.0339	18.8528	18.3816	17.5777	18.0802	17.9807
MAD	9.6586	8.4361	7.6692	6.6087	4.2507	3.1007	2.5924
MLE	15.3643	16.3107	16.4671	16.9253	17.4719	17.7281	17.8609
MAD	7.1586	6.2665	5.2678	4.6483	3.3252	2.2616	1.5662

6.1. *TME and MLE for various sample sizes*

Table 1 contains the simulated estimates for various sample sizes. These results show that the MLE tends to overestimate and the TME tends to underestimate the slope parameter β_2, whereas this is converse for the intercept β_1 and the variance σ_u. The TME performs rather satisfactory for the sample size $T \geqslant 40$ and so does the MLE for $T \geqslant 50$. The MLE seems to underestimate σ_u significantly. However, for $T \geqslant 100$ the estimates of both procedures are rather satisfactory with the precision 0.0097 for β_1 and β_2 (except the MLE by $T = 200$, which is 0.0324). For the sample size $T \leqslant 30$, both procedures may have big bias for certain parameters, which is not surprising if we note that only about 23 observations were used to estimate five parameters. Note also, in general, β_1 and β_2 are more exactly estimated than σ_u. Finally, the MLE has almost always smaller MAD than the TME. However, this is a trade-off for the fact that the MLE is computationally much more expensive.

6.2. *Sensitivities of TME and MLE to misspecified Δ*

Table 2 shows the results of simulations for various Δ's, whereas the true $\Delta = 0.1$. For easier comparison the estimates for the true $\Delta = 0.1$ are also included. Note that the case $\Delta = 0.001$ approximates the error-free modeling, whereas the data are, in fact, generated from an errors-in-variables system. In order to reduce

Table 2
TME and MLE for $T = 200$ and true $\Delta = 0.1$

Δ	0.001	0.01	0.05	0.1	0.15	0.2	0.3
$\beta_1 = -6$							
TME	−4.9198	−4.9906	−5.4024	−6.0078	−6.5174	−7.0931	−8.1172
MAD	1.2381	1.1759	0.9568	0.9007	0.9418	1.2862	2.1372
MLE	−4.9203	−5.0199	−5.4760	−6.0324	−6.5351	−7.1070	−8.1951
MAD	1.1530	1.0787	0.8204	0.6881	0.8294	1.2093	2.2020
$\beta_2 = 0.6$							
TME	0.5443	0.5498	0.5689	0.5992	0.6254	0.6530	0.7050
MAD	0.0585	0.0538	0.0406	0.0343	0.0390	0.0581	0.1052
MLE	0.5453	0.5517	0.5723	0.6013	0.6270	0.6549	0.7010
MAD	0.0558	0.0502	0.0356	0.0272	0.0354	0.0571	0.1090
$\sigma_u = 18$							
TME	23.7936	23.2596	20.8251	18.0802	15.0082	12.5844	6.7985
MAD	6.1017	5.6881	4.1465	3.1007	4.3896	5.9605	11.2415
MLE	23.5279	22.9712	20.6406	17.7281	14.6589	12.0428	5.7226
MAD	5.5572	4.9975	3.0457	2.2616	3.7282	6.0374	12.2774

the sampling effect, in each simulation $T = 200$ observations are generated. From Table 2 we see that, first of all, both the TME and the MLE are rather sensitive to changes of Δ and their sensitivities have a very clear systematic pattern described by (7) in Section 3. The MLE almost always overestimate the TME for β_2 and underestimate the TME for β_1 and σ_u. Different from Section 6.1, the MLE have smaller MAD than TME do for $\Delta \leqslant 0.2$ (except for σ_u and $\Delta = 0.2$) and larger MAD for $\Delta \geqslant 0.3$. In general, the MLE seem to be more sensitive than the TME are.

7. Conclusions

We have shown that if the covariates in a censored regression model are measured with errors, then the problem of non-identifiability occurs as in the usual linear errors-in-variables models under normality. We proposed two-step moment estimators (TME) of the model and derived the asymptotic covariance matrices. The obtained formulae can be used for the moment estimator of the usual error-free Tobit model. We also demonstrated that, given the normality assumption and the identifying condition, this model can be uniquely reduced to an error-free Tobit model. As a result, estimators for the original model may be obtained via the estimators for the reduced model. In particular, the maximum-likelihood estimator (MLE) is derived in this way. The MLE may be calculated by standard numerical methods because of the uniqueness of the global maximum, whereas the TME have the practical advantage that the numerical calculation is very easy. The Monte-Carlo studies show that, in general, both procedures produce rather satisfactory estimates for sample size larger than 50 for a simple model with five unknown parameters. However, they are rather sensitive to the misspecification of the a priori information used. Thus, the specification of the a priori information is important in applications. In general, the TME is more stable and reliable than the MLE and also computationally cheaper.

Appendix A Proof of (10)–(13)

For notational simplicity, we denote $\Phi=\Phi(\delta)$, $\phi=\phi(\delta)$, $\lambda=\phi/\Phi$ and, correspondingly, $\hat{\Phi}=\Phi(\hat{\delta})$, $\hat{\phi}=\phi(\hat{\delta})$ and $\hat{\lambda}=\hat{\phi}/\hat{\Phi}$. For various matrix operations and rules used in this paper we refer the reader to Magnus and Neudecker (1988) and Lütkepohl (1991).

First we show (10). By (3) and straightforward calculation, we have

$$\begin{aligned}\hat{\mu}_\eta-\mu_\eta &= \frac{\hat{\delta}\hat{\mu}_{y+}}{\hat{\delta}+\hat{\lambda}}-\frac{\delta\mu_{y+}}{\delta+\lambda}\\ &= \frac{\hat{\delta}(\hat{\mu}_{y+}-\mu_{y+})}{\hat{\delta}+\hat{\lambda}}+\frac{\mu_{y+}[\lambda(\hat{\delta}-\delta)-\delta(\hat{\lambda}-\lambda)]}{(\hat{\delta}+\hat{\lambda})(\delta+\lambda)}\\ &= \frac{\hat{\delta}(\hat{\mu}_{y+}-\mu_{y+})}{\hat{\delta}+\hat{\lambda}}+\frac{\mu_{y+}[\lambda-\delta\lambda_\delta(\tilde{\delta})](\hat{\delta}-\delta)}{(\hat{\delta}+\hat{\lambda})(\delta+\lambda)},\end{aligned}\tag{A.1}$$

where $\lambda_\delta(\tilde{\delta})=-\lambda(\tilde{\delta})[\tilde{\delta}+\lambda(\tilde{\delta})]$ is the derivative of $\lambda(\delta)$ and $\tilde{\delta}$ lies between $\hat{\delta}$ and δ. By definition, we have

$$\begin{aligned}\hat{\delta}-\delta &= \frac{1}{\phi(\tilde{\delta})}\left(\frac{\hat{\mu}_y}{\hat{\mu}_{y+}}-\Phi\right)\\ &= \frac{1}{\phi(\tilde{\delta})}\left(\frac{1}{T}\sum_{t=1}^{T}s_t-\Phi\right).\end{aligned}\tag{A.2}$$

It follows from (A.1) and (A.2) that

$$\begin{aligned}\hat{\mu}_\eta-\mu_\eta &= \frac{\mu_\eta[\lambda-\delta\lambda_\delta(\tilde{\delta})]}{\delta(\hat{\delta}+\hat{\lambda})\phi(\tilde{\delta})}\left(\frac{1}{T}\sum_{t=1}^{T}s_t-\Phi\right)+\frac{\hat{\mu}_\eta}{\hat{\mu}_{y+}}(\hat{\mu}_{y+}-\mu_{y+})\\ &= \hat{a}(\bar{s}-\Phi)+\hat{b}(\hat{\mu}_{y+}-\mu_{y+}).\end{aligned}$$

Similarly,

$$\begin{aligned}
\hat{\sigma}_\eta - \sigma_\eta &= \frac{\hat{\mu}_\eta^2}{\hat{\delta}^2} - \frac{\mu_\eta^2}{\delta^2} \\
&= \frac{\hat{\mu}_\eta + \mu_\eta}{\hat{\delta}^2}(\hat{\mu}_\eta - \mu_\eta) - \frac{\mu_\eta^2(\hat{\delta}+\delta)}{\hat{\delta}^2\delta^2}(\hat{\delta} - \delta) \\
&= \left[\frac{(\hat{\mu}_\eta + \mu_\eta)\hat{a}}{\hat{\delta}^2} - \frac{\sigma_\eta(\hat{\delta}+\delta)}{\hat{\delta}^2\phi(\tilde{\delta})}\right](\bar{s} - \Phi) + \frac{(\hat{\mu}_\eta + \mu_\eta)\hat{b}}{\hat{\delta}^2}(\hat{\mu}_{y+} - \mu_{y+}) \\
&= \hat{c}(\bar{s} - \Phi) + \hat{d}(\hat{\mu}_{y+} - \mu_{y+}).
\end{aligned}$$

Furthermore, we have

$$\begin{aligned}
\hat{\sigma}_{x\eta} - \sigma_{x\eta} &= (\hat{\mu}_{xy+} - \mu_{xy+}) - \hat{\mu}_x(\hat{\mu}_{y+} - \mu_{y+}) - \mu_{y+}(\hat{\mu}_x - \mu_x) \\
&= -\hat{\mu}_x\left(\frac{T}{T_1}\right)\frac{1}{T}\sum_{t=1}^{T} s_t(y_t - \mu_{y+}) - \mu_{y+}(\hat{\mu}_x - \mu_x) \\
&\quad + \left(\frac{T}{T_1}\right)\frac{1}{T}\sum_{t=1}^{T} s_t(x_t y_t - \mu_{xy+})
\end{aligned}$$

and

$$\hat{\sigma}_x - \sigma_x = (\hat{\mu}_{x \otimes x} - \mu_{x \otimes x}) - (I_k \otimes \mu_x + \hat{\mu}_x \otimes I_k)(\hat{\mu}_x - \mu_x).$$

Putting these equations into matrix form we obtain (10) with

$$\hat{A} = \begin{pmatrix}
\hat{a} & (T/T_1)\hat{b} & 0 & 0 & 0 \\
\hat{c} & (T/T_1)\hat{d} & 0 & 0 & 0 \\
0 & -(T/T_1)\hat{\mu}_x & (T/T_1)I_k & -\mu_{y+}I_k & 0 \\
0 & 0 & 0 & I_k & 0 \\
0 & 0 & 0 & -I_k \otimes \mu_x - \hat{\mu}_x \otimes I_k & I_k
\end{pmatrix}.$$

Using relationships (3) it is straightforward to have

$$\begin{aligned}
\operatorname*{plim}_{T\to\infty} \hat{a} &= \frac{\mu_\eta[\lambda + \delta\lambda(\delta+\lambda)]}{\delta(\delta+\lambda)\phi(\delta)} = \frac{\mu_\eta[1 + \delta(\delta+\lambda)]}{\delta(\delta+\lambda)\Phi} \\
&= \frac{\sigma_\eta + \mu_\eta\mu_{y+}}{\mu_y} = \frac{\mu_{y^2+}}{\mu_y},
\end{aligned}$$

$$\operatorname*{plim}_{T\to\infty} \hat{b} = \frac{\mu_\eta}{\mu_{y+}},$$

$$\operatorname*{plim}_{T\to\infty} \hat{c} = \frac{2\mu_\eta a}{\delta^2} - \frac{2\sigma_\eta}{\delta\phi(\delta)} = 2\sigma_\eta\left(\frac{\mu_{y^2+}}{\mu_\eta\mu_y} - \frac{1}{\delta\phi(\delta)}\right)$$

and

$$\operatorname*{plim}_{T\to\infty} \hat{d} = \frac{2\mu_\eta b}{\delta^2} = \frac{2\sigma_\eta}{\mu_{y+}}.$$

Finally, by definition,

$$\operatorname*{plim}_{T\to\infty}\left(\frac{T_1}{T}\right) = \operatorname*{plim}_{T\to\infty} \frac{\hat{\mu}_y}{\hat{\mu}_{y+}} = \Phi.$$

It follows that $\operatorname{plim}_{T\to\infty} \hat{A} = A$ which is given in (12).

Next we show (11). Using (9) we have

$$\hat{\beta}_2 - \beta_2 = \hat{\Sigma}_\xi^{-1}\hat{\sigma}_{x\eta} - \Sigma_\xi^{-1}\sigma_{x\eta}$$

$$= \hat{\Sigma}_\xi^{-1}(\hat{\sigma}_{x\eta} - \sigma_{x\eta}) - \hat{\Sigma}_\xi^{-1}(\hat{\Sigma}_x - \Sigma_x)\Sigma_x^{-1}\sigma_{x\eta}$$

$$= \hat{\Sigma}_\xi^{-1}(\hat{\sigma}_{x\eta} - \sigma_{x\eta}) - (\sigma_{x\eta}'\Sigma_x^{-1} \otimes \hat{\Sigma}_\xi^{-1})(\hat{\sigma}_x - \sigma_x),$$

$$\hat{\beta}_1 - \beta_1 = (\hat{\mu}_\eta - \mu_\eta) - \hat{\beta}_2'(\hat{\mu}_\xi - \mu_\xi) - (\hat{\beta}_2 - \beta_2)'\mu_\xi$$

$$= (\hat{\mu}_\eta - \mu_\eta) - \hat{\beta}_2'(\hat{\mu}_x - \mu_x) - \mu_x'\hat{\Sigma}_\xi^{-1}(\hat{\sigma}_{x\eta} - \sigma_{x\eta})$$

$$+ \mu_x'(\sigma_{x\eta}'\Sigma_x^{-1} \otimes \hat{\Sigma}_\xi^{-1})(\hat{\sigma}_x - \sigma_x),$$

$$\hat{\sigma}_u - \sigma_u = (\hat{\sigma}_\eta - \sigma_\eta) - \hat{\beta}_2'(\hat{\sigma}_{x\eta} - \sigma_{x\eta}) - (\hat{\beta}_2 - \beta_2)'\sigma_{x\eta}$$

$$= (\hat{\sigma}_\eta - \sigma_\eta) - (\hat{\beta}_2' + \sigma_{x\eta}'\hat{\Sigma}_\xi^{-1})(\hat{\sigma}_{x\eta} - \sigma_{x\eta})$$

$$+ \sigma_{x\eta}'(\sigma_{x\eta}'\Sigma_x^{-1} \otimes \hat{\Sigma}_\xi^{-1})(\hat{\sigma}_x - \sigma_x),$$

$$\hat{\mu}_\xi - \mu_\xi = \hat{\mu}_x - \mu_x$$

and

$$\hat{\sigma}_{\xi} - \sigma_{\xi} = \mathrm{vec}\,(\hat{\Sigma}_{\xi} - \Sigma_{\xi})$$

$$= \mathrm{vec}\,[(\hat{\Sigma}_x - \Sigma_x)(I + \Delta)^{-1}]$$

$$= [\Sigma_{\xi}\Sigma_x^{-1} \otimes I_k](\hat{\sigma}_x - \sigma_x).$$

Then (11) follows with

$$\hat{B} = \begin{pmatrix} 1 & 0 & -\mu_x'\hat{\Sigma}_{\xi}^{-1} & -\hat{\beta}_2' & \sigma_{x\eta}'\Sigma_x^{-1} \otimes \mu_x'\hat{\Sigma}_{\xi}^{-1} \\ 0 & 0 & \hat{\Sigma}_{\xi}^{-1} & 0 & -\sigma_{x\eta}'\Sigma_x^{-1} \otimes \hat{\Sigma}_{\xi}^{-1} \\ 0 & 1 & -\hat{\beta}_2' - \sigma_{x\eta}'\hat{\Sigma}_{\xi}^{-1} & 0 & \sigma_{x\eta}'\Sigma_x^{-1} \otimes \sigma_{x\eta}'\hat{\Sigma}_{\xi}^{-1} \\ 0 & 0 & 0 & I_k & 0 \\ 0 & 0 & 0 & 0 & \Sigma_{\xi}\Sigma_x^{-1} \otimes I_k \end{pmatrix}$$

which converges in probability to the matrix B in (13). The proof is completed. □

References

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

Amemiya, T., 1984. Tobit models: a survey. Journal of Econometrics 24, 3–61.

Amemiya, T., 1985. Advanced Econometrics. Basil Blackwell, Oxford, UK.

Colombi, R., 1993. Stochastic frontier and switching regression models with latent variables. In: Haagen, K., Bartholomew, D., Deistler, M. (Eds.), Statistical Modeling and Latent Variables. North-Holland, Amsterdam.

Fuller, W.A., 1987. Measurement Error Models. Wiley, New York.

Gleser, L.J., 1992. The importance of assessing measurement reliability in multivariate regression. Journal of the American Statistical Association 87, 696–707.

Greene, W.H., 1983. Estimation of limited dependent variable models by ordinary least squares and the method of moments. Journal of Econometrics 21, 195–212.

Greene, W.H., 1993. Econometric Analysis, 2nd ed. Macmillan, New York.

Heckman, J.J., MaCurdy, T.E., 1986. Labor econometrics. In: Griliches, Z., Intriligator, M.D. (Eds.), Handbook of Econometrics, vol. III. North-Holland, Amsterdam.

Hsiao, C., 1983. Identification. In: Griliches, Z., Intriligator, M.D. (Eds.), Handbook of Econometrics, vol. I, North-Holland, Amsterdam.

Hsiao, C., 1991. Identification and estimation of dichotomous latent variables models using panel data. Review of Economic Studies 58, 717–731.

Jaech, J.L., 1985. Statistical Analysis of Measurement Errors. Wiley, New York.

Killingsworth, M.R., Heckman, J.J., 1986. Female labor supply: a survey. In: Ashenfelter, O., Layard, R. (Eds.), Handbook of Labor Economics, vol. I. North-Holland, Amsterdam.

Lütkepohl, H., 1991. Introduction to Multiple Time Series Analysis. Springer, Berlin.

Maddala, G.S., 1985. Limited-dependent and qualitative variables in econometrics. Cambridge University Press, Cambridge.

Magnus, J.R., Neudecker, H., 1988. Matrix Differential Calculus with Applications in Statistics and Econometrics. Wiley, New York.

Olsen, R.J., 1978. Note on the uniqueness of the maximum likelihood estimator for the Tobit model. Econometrica 46, 1211–1215.

Pencavel, J., 1986. Labor supply of men: a survey. In: Ashenfelter, O., Layard, R. (Eds.), Handbook of Labor Economics, vol. I. North-Holland, Amsterdam.

Wang, L., 1994. Identification and estimation of censored regression model with errors-in-variables. USC-Economics Arthur Andersen Working Paper Series #9401, University of Southern California, Los Angeles.

Weiss, A.A., 1993. Some aspects of measurement error in a censored regression model. Journal of Econometrics 56, 169–188.