IDENTIFIABILITY AND ESTIMATION OF CENSORED ERRORS-IN-VARIABLES MODELS *

Liqun Wang, University of Basel, Switzerland
University of Southern California, Los Angeles, USA

KEY WORDS: Censored regression model, Tobit model, measurement error, identification, moment estimator, maximum likelihood.

1 Introduction

The censored regression (Tobit) model has been widely applied in econometrics, biometrics, psychometrics and many other fields (Amemiya (1985)). Usually the regressors in these models are assumed to be non-random bounded constants. This assumption is obviously not always appropriate in many situations and may lead to inaccurate and inconsistent estimates.

Recently a class gf errors-in-variables (EV) models with binary dependent variables are studied by Hsiao (1991) and the least absolute deviation estimators of the models with censored dependent variables are investigated by Weiss (1993). In this paper we consider the following censored (linear) errors-in-variables (CEV) model

$$\begin{aligned} \eta_t &= \beta_1 + \beta_2'\xi_t + u_t, \\ y_t &= \max\{\eta_t, 0\}, \\ x_t &= \xi_t + v_t, \end{aligned} \tag{1.1}$$

where $\eta_t \in I\!R$, $\xi_t \in I\!R^k$ are the unobserved variables, y_t, x_t the observed variables, u_t, v_t the errors and β_1, β_2 the regression parameters. Furthermore we assume that u_t, v_t and ξ_t be independently and normally distributed $(u_t, v_t', \xi_t')' \sim N\left[(0,0,\mu_\xi')', diag(\sigma_u, \Sigma_v, \Sigma_\xi)\right]$, where $\mu_\xi \in I\!R^k$, $\sigma_u > 0$, Σ_v and Σ_ξ are $k \times k$ non-negative definite matrices.

The major difference between model (1.1) and the Tobit model is that the independent variable ξ_t is not exactly observed and, as a result, x_t's are no longer constants and the distributions of x_t's enter the likelihood function of the model. This feature arises many difficulties and complexities in conducting the statistical analysis of the model. Another feature of model (1.1) is that, as the usual linear EV model, it suffers from the problem of non-identifiability in general, as is shown in section 2. In section 3, under the condition of given $\Sigma_\xi^{-1}\Sigma_v$, a two-step moment estimation procedure (TME) is proposed which is consistent and asymptotically normal. Under the same condition the maximum likelihood estimators (MLE) are treated in section 4. It is shown that a suitably reparametrized likelihood function is globally concave and therefore the unique global MLE exist. The consistency and asymptotic normality of the MLE are also shown.

*I would like to thank Shoaleh Ghashghai, Klaus Poetzelberger and Wolfgang Polasek for helpful comments. Financial support from Swiss National Science Foundation is gratefully acknowledged.

2 Identifiability

For identifiability we adopt the definition of Hsiao (1983) or Fuller (1987). Formally, let z be the vector of $(y_t, x_t')'$, $t = 1, 2, ..., T$, where T is the sample size, and suppose the sample distribution function $F(z|\theta)$ be known up to an n-dimensional unknown parameter vector θ. Let $\Theta \subseteq I\!R^n$ be the natural parameter space. Then the model is said to be identifiable, if for any $\theta_1, \theta_2 \in \Theta$, $F(z|\theta_1) \equiv F(z|\theta_2)$ 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 the variable y_t has a so-called censored normal distribution, i.e., y_t takes the value 0 with probability $P(\eta_t \le 0) = \Phi(-\mu_\eta/\sqrt{\sigma_\eta})$ and has the density function $f(y_t) = (1/\sqrt{\sigma_\eta})\phi[(y_t - \mu_\eta)/\sqrt{\sigma_\eta}]$ for $y_t > 0$, where $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal distribution and density functions. Thus the joint distribution of $(y_t, x_t')'$ is either normal or censored normal or the mixture of both. In any case it is uniquely determined by the first two moments and the conditional moments of $(y_t, x_t')'$ given $y_t > 0$. However, it may be easily verified that all these moments and conditional moments are uniquely determined by the first two moments of $(\eta_t, x_t')'$ and vice versa (Wang (1993)), as is easily seen by the equations

$$E(y_t|y_t > 0) = \mu_\eta + \sqrt{\sigma_\eta}\lambda(\gamma), \tag{2.1}$$

$$E(y_t^2|y_t > 0) = \sigma_\eta + \mu_\eta E(y_t|y_t > 0), \tag{2.2}$$

$$E(x_t y_t|y_t > 0) = \sigma_{x\eta} + \mu_x E(y_t|y_t > 0), \tag{2.3}$$

where $\gamma = \mu_\eta/\sqrt{\sigma_\eta}$ and $\lambda(\cdot) = \phi(\cdot)/\Phi(\cdot)$. This means that the first two moments of $(\eta_t, x_t')'$ contain all observational information from data. Therefore the problem of identifiability is reduced to, first, writing down all first two moments of $(\eta_t, x_t')'$ which are the functions (defined by the model) of unknown parameters, and then, inspecting whether all parameters are uniquely determined by the moment equations. As for the usual linear EV model, we have

$$\mu_x = \mu_\xi, \qquad \Sigma_x = \Sigma_\xi + \Sigma_v, \tag{2.4}$$

and

$$\begin{aligned} \mu_\eta &= \beta_1 + \beta_2'\mu_\xi, \\ \sigma_\eta &= \beta_2'\Sigma_\xi\beta_2 + \sigma_u, \\ \sigma_{x\eta} &= \Sigma_\xi\beta_2. \end{aligned} \tag{2.5}$$

It is easily seen that in (2.4) – (2.5) the number of free parameters exceeds the number of equations and hence, except μ_ξ, all other parameters on the right-hand side are not identified.

Theorem 2.1 *In model (1.1) only the parameters μ_ξ, μ_η, σ_η and $\sigma_{x\eta}$ are identified whereas the parameters β_1, β_2, σ_u, Σ_v and Σ_ξ are not identified. Thus in general model (1.1) is not identifiable.*

According to Theorem 2.1 in order to obtain the unique estimates of model (1.1) a certain kind a priori information is needed. For example, in many real problems the variance ratio $\sigma_u^{-1}\Sigma_v$ may be estimated by repeated sampling or determined from some sources independent of data. In some other cases the information about the reliability ratio $\kappa = \Sigma_x^{-1}\Sigma_\xi$ may be available (Gleser (1992)). Since $\kappa = (I + \Sigma_\pi^{-1}\Sigma_v)^{-1}$, this is equivalent to the assumption that the noise-to-signal ratio $\Delta = \Sigma_\xi^{-1}\Sigma_v$ is known. In the following two sections we discuss two estimation procedures for the given Δ.

3 Two-step Moment Estimators (TME)

Suppose the data $(y_t, x_t)'$, $t = 1, 2, ..., T$ be given, in which T_0 y_t's are zero and $T_1 = T - T_0$ y_t's are positive. Without loss of generality we assume $0 < T_0 < T$. Then the conditional moments on the left-hand side of (2.1) – (2.3) are consistently estimated by the sample moments using the positive y_t's and the corresponding x_t's. Using the analogous notation we denote these estimators as $\hat\mu_{y|y>0}$, $\hat\mu_{y^2|y>0}$ and $\hat\mu_{xy|y>0}$. Since $E(y_t) = E(y_t|y_t > 0)\Phi(\gamma)$, we may estimate γ by $\hat\gamma = \Phi^{-1}(\bar y/\hat\mu_{y|y>0})$, where $\bar y = (1/T)\sum_{t=1}^T y_t$, then by (2.1) – (2.3), σ_η, μ_η and $\sigma_{x\eta}$ are estimated as

$$\begin{aligned} \hat\sigma_\eta &= \hat\mu_{y^2|y>0}/(1 + \hat\gamma^2 + \hat\gamma\lambda(\hat\gamma)), \\ \hat\mu_\eta &= \hat\gamma\sqrt{\hat\sigma_\eta}, \\ \hat\sigma_{x\eta} &= \hat\mu_{xy|y>0} - \bar x\hat\mu_{y|y>0}, \end{aligned} \tag{3.1}$$

where, $\bar x = (1/T)\sum_{t=1}^T x_t$. Now substituting σ_η, μ_η and $\sigma_{x\eta}$ in the equations (2.4) – (2.5) through the estimates (3.1) and solving these equations we obtain

$$\begin{aligned} \hat\mu_\xi &= \bar x, \\ \hat\Sigma_\xi &= S_x(I + \Delta)^{-1}, \\ \hat\Sigma_v &= \hat\Sigma_\xi\Delta, \end{aligned} \tag{3.2}$$

and

$$\begin{aligned} \hat\beta_2 &= \hat\Sigma_\xi^{-1}\hat\sigma_{x\eta}, \\ \hat\beta_1 &= \hat\mu_\eta - \hat\beta_2'\hat\mu_\xi, \\ \hat\sigma_u &= \hat\sigma_\eta - \hat\beta_2'\hat\sigma_{x\eta}, \end{aligned} \tag{3.3}$$

where $S_x = (1/T)\sum_{t=1}^T (x_t - \bar x)(x_t - \bar x)'$.

The consistency and asymptotic normality of the estimators in (3.1) – (3.2) follow immediately from that of the sample moments. The asymptotic properties for the estimators in (3.3) are given in the following theorem, the proof of which is given in Wang (1993).

Theorem 3.1 *Let $\theta = (\beta_1, \beta_2', \sigma_u)'$ and $\hat\theta_{TM}$ the corresponding TME. If $\Sigma_\xi > 0$ and $\Delta = \Sigma_\xi^{-1}\Sigma_v$ is given, then*

1. *$\hat\theta_{TM} \overset{a.s.}{\to} \theta_0$, as $T \to \infty$, where θ_0 are the true parameters of model* (1.1).

2. *$\sqrt{T}(\hat\theta_{TM} - \theta_0) \overset{L}{\to} N\big(0, \Sigma(\theta_0)\big)$, where the symbol $\overset{L}{\to}$ denotes the convergence in law and $\Sigma(\theta_0)$ is the covariance matrix with the components defined by*
$\Sigma_{11} = \sigma_u + \beta_2'\Sigma_v\beta_2 + \mu_x'\Sigma_{22}\mu_x$,
$\Sigma_{12} = -\mu_x'\Sigma_{22}$,
$\Sigma_{13} = -\mu_x'\Sigma_{23}$,
$\Sigma_{22} = (\sigma_\eta - \sigma_{x\eta}'\Sigma_x^{-1}\sigma_{x\eta})(I + \Delta)\Sigma_\xi^{-1}$,
$\Sigma_{23} = -2(\sigma_\eta - \sigma_{x\eta}'\Sigma_x^{-1}\sigma_{x\eta})\Delta\beta_2$,
$\Sigma_{33} = 2\sigma_u^2 + \beta_2'[4\sigma_\eta\Sigma_v + \sigma_{x\eta}'\Sigma_x^{-1}\sigma_{x\eta}(4\Sigma_v + \Sigma_x) + 7\sigma_{x\eta}\sigma_{x\eta}']\beta_2$.

Remark 3.1 The two-step procedure in this section may be similarly applied to the case where instead of the noise-to-signal ratio $\Sigma_\xi^{-1}\Sigma_v$ 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 properties of the estimators may be established analogously to Theorem 3.1. Such results for the univariate model ($k = 1$) are given by Theorem 1.3.1 of Fuller (1987), page 32.

4 Maximum Likelihood Estimators (MLE)

In this section we consider the maximum likelihood estimators of model (1.1). Let the data be given as in last section. Since $(\eta_t, x_t')'$ are jointly normal, the conditional distribution of η_t given x_t is again normal with conditional mean $E(\eta_t|x_t)$ and conditional variance $V(\eta_t|x_t)$. Note that $V(\eta_t|x_t)$ does not depend on the subscript t and hence will be written as $V(\eta|x)$. As $y_t = \eta_t$ when $y_t > 0$, the log-likelihood function of model (1.1) is, up to a constant,

$$\begin{aligned} L = & \sum_0 \log\Phi\Big(-\frac{E(\eta_t|x_t)}{\sqrt{V(\eta|x)}}\Big) - \frac{T_1}{2}\log V(\eta|x) \\ & -\frac{1}{2V(\eta|x)}\sum_1 [y_t - E(\eta_t|x_t)]^2 - \frac{T}{2}\log|\Sigma_x| \\ & -\frac{1}{2}\sum (x_t-\mu_x)'\Sigma_x^{-1}(x_t-\mu_x), \end{aligned} \tag{4.1}$$

where μ_η, μ_x, σ_η, $\sigma_{x\eta}$ and Σ_x are given by (2.4) – (2.5). Throughout this paper we use $\sum_0$ to denote the summation over the t's for which $y_t = 0$, $\sum_1$ over the t's for which $y_t > 0$ and $\sum$ without index to denote the summation over all observations.

Again because of the problem of non-identifiability the likelihood function (4.1) does not have finite maximum without further condition. In order to guarantee the existence of the finite maximum we make again the assumption that $\Delta = \Sigma_\xi^{-1}\Sigma_v$ be known. Thus the free parameters in (4.1) are $(\beta_1, \beta_2', \sigma_u, \mu_\xi', \text{vech}\Sigma_\xi')'$, where vech is the vectorization operator such that vechΣ is a vector consisting of main and lower diagonal elements of Σ. Let the natural parameter space be denoted as $\tilde{\Theta}$. As in the non-EV case, directly working with the likelihood function (4.1) turns out to be cumbersome. In order to simplify (4.1) we define the reparametrization $\mathcal{T}$: $\tilde{\Theta} \mapsto \tilde{\Theta}$ as

$$\mu_x = \mu_\xi, \qquad \Sigma_x = \Sigma_\xi(I+\Delta), \tag{4.2}$$

and

$$\begin{aligned} \tau &= 1/\sqrt{V(\eta|x)}, \\ \alpha_1 &= \tau\beta_1, \\ \alpha_2 &= \tau\Sigma_x^{-1}\Sigma_\xi\beta_2. \end{aligned} \tag{4.3}$$

Then the new parameters are $(\alpha_1, \alpha_2', \tau, \mu_x', \text{vech}\Sigma_x')'$. Obviously $\mathcal{T}$ is a homeomorphism from $\tilde{\Theta}$ onto $\tilde{\Theta}$. The log-likelihood function may be further simplified by noting that, since only the variable y_t is censored, the MLE of μ_x and Σ_x should not be affected by the censoring and therefore should be the same as in the non-censored linear model. Indeed, setting the first derivatives of L to zero we may obtain the MLE for μ_x and Σ_x which are given by $\hat{\mu}_x = \bar{x}$ and $\hat{\Sigma}_x = S_x$. Inserting these estimates into L results in the concentrated log-likelihood function which has only $k+2$ unknown parameters $\psi = (\alpha_1, \alpha_2', \tau)'$ and is defined on the parameter space $\Theta = \mathbb{R}^{k+1} \times \mathbb{R}_+$.

Without loss of generality we assume that the first T_0 y_t be zero and let Y_1 be the vector containing the T_1 positive y_t's. Furthermore denote $\tilde{x}_t = (1, (x_t + \Delta'\bar{x})')'$, $X_0 = (\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_{T_0})'$, $X_1 = (\tilde{x}_{T_0+1}, \tilde{x}_{T_0+2}, \cdots, \tilde{x}_T)'$ and $Z = (X_1, -Y_1)$. Then after some reordering and arrangements of terms the concentrated log-likelihood function may be written up to a constant as

$$L_c(\psi) = \sum_0 \log\Phi(-\alpha'\tilde{x}_t) - \frac{1}{2}\psi' Z' Z\psi + T_1\log\tau,$$

where $\alpha = (\alpha_1, \alpha_2')'$. In order to show that $L_c(\psi)$ has unique maximum we calculate its second derivatives

$$\frac{\partial^2 L_c(\psi)}{\partial\psi\partial\psi'} = -\begin{pmatrix} X_0'\Lambda_0 X_0 & 0 \\ 0 & T_1/\tau^2 \end{pmatrix} - Z'Z, \tag{4.4}$$

where $\lambda_0 = (\lambda_t, t = 1, 2, ..., T_0)'$, $\lambda_t = \phi(\alpha'\tilde{x}_t)/\Phi(-\alpha'\tilde{x}_t)$ and $\Lambda_0 = \text{diag}[\lambda_t(\lambda_t - \alpha'\tilde{x}_t), t = 1, 2, ..., T_0]$. Since $\lambda_t(\lambda_t - \alpha'\tilde{x}_t) > 0$, the Hessian matrix (4.5) is negative definite with probability one for all $\psi \in \Theta$, which means $L_c(\psi)$ is concave in the space Θ with probability one. Furthermore, it is easy to verify that the gradient of $L_c(\psi)$ will change sign when α tends to positive infinity and to negative infinity, which means that $L_c(\psi)$ attains the maximum at an interior point Θ. Since the mapping $\mathcal{T}$ is continuous, by the invariance principle of the MLE we have the following results.

Theorem 4.1 *Suppose $\sigma_u > 0$, $\Sigma_\xi > 0$ and $\Delta = \Sigma_\xi^{-1}\Sigma_v$ be given. If the Hessian matrix (4.4) is nonsingular, then*

1. *The concentrated log-likelihood function $L_c(\psi)$ is concave in Θ and hence has a unique, finite, global maximum.*

2. *The likelihood function L in (4.1) has a unique, finite, global maximum in the parameter space $\tilde{\Theta}$.*

Let $\hat{\psi}$ be the value obtained by maximizing $L_c(\psi)$, then the MLE for the original parameters $(\beta_1, \beta_2', \sigma_u)'$ are calculated according to (4.3). The MLE for μ_ξ, Σ_ξ and Σ_v are calculated by (4.2) and therefore are identical with the TME (3.2).

Now we consider the asymptotic properties of the MLE so obtained. As for the TME we consider only the MLE of $\theta = (\beta_1, \beta_2', \sigma_u)'$, which is denoted by $\hat{\theta}_{ML}$. First we have the following theorem (Wang (1993)).

Theorem 4.2 *Let ψ_0 be the value corresponding to the true parameters of model* (1.1). *Then under the conditions of Theorem 4.1, it holds*

1. $\hat{\psi} \xrightarrow{P} \psi_0$, *as* $T \to \infty$.

2. $\sqrt{T}(\hat{\psi} - \psi_0) \xrightarrow{L} N\big(0, -H(\psi_0)^{-1}\big)$, *where*

$$H(\psi_0) = \operatorname*{plim}_{T\to\infty} \frac{1}{T} \frac{\partial^2 L_c(\psi_0)}{\partial\psi\partial\psi'}.$$

Now the consistency of $\hat{\theta}_{ML}$ follows immediately from Theorem 4.2 and the continuity of the mapping $\mathcal{T}$. To show the asymptotic normality of $\hat{\theta}_{ML}$, let $\theta(\psi) : \Theta \mapsto \Theta$ denote the mapping induced by $\mathcal{T}$. Then $\theta(\psi)$ is continuously differentiable and hence we have the first-order Taylor expansion

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

where $\tilde{\psi}$ lies between $\hat{\psi}$ and ψ and

$$\frac{\partial\theta}{\partial\psi'} = \frac{1}{\sqrt{V(\eta|x)}} \begin{pmatrix} 1 & 0 & -\beta_1 \\ 0 & I+\Delta & -\beta_2 \\ 0 & -2\Sigma_v\beta_2 & -2\sigma_u \end{pmatrix}$$

The derivative $\partial\theta/\partial\psi'$ is obviously a continuous function of ψ. Thus by Theorem 4.2 we have the following results.

Theorem 4.3 *Let θ_0 be the true parameters of model* (1.1). *Then under the conditions of Theorem 4.1, $\hat{\theta}_{ML}$ satisfy*

1. *Consistency:* $\hat{\theta}_{ML} \xrightarrow{P} \theta_0$.

2. *Asymptotic normality:*

$$\sqrt{T}(\hat{\theta}_{ML} - \theta_0) \xrightarrow{L} N\big(0, \Sigma(\theta_0)\big),$$

where

$$\Sigma(\theta_0) = \frac{\partial\theta(\psi_0)}{\partial\psi'} H(\psi_0)^{-1} \Big(\frac{\partial\theta(\psi_0)}{\partial\psi'}\Big)'$$

is evaluated at θ_0.

Remark 4.1 The maximization of the function $L_c(\psi)$ may be carried out through standard numerical methods such as Newton-Raphson. Since $L_c(\psi)$ is globally concave, the iteration may start at any finite point. However for a rapid convergence a good starting point is important. The TME derived in section 3 may serve as initial values for starting the iteration. As is shown by a Monte Carlo simulation in Wang (1993), in an univariate model the MLE procedures using Newton-Raphson algorithm and the TME as starting values may achieve rather satisfactory convergence after four or five iterations. Furthermore, it is well-known that the estimators $\hat{\psi}_1$ and $\hat{\theta}_1$ obtained after one iteration in Newton-Raphson procedure have the same asymptotic distributions as the MLE $\hat{\psi}$ and $\hat{\theta}_{ML}$ respectively.

References

[1] Amemiya, T. (1985). Advanced Econometrics. Basil Blackwell, Oxford, UK.

[2] Fuller, W.A. (1987). Measurement Error Models. Wiley, New York.

[3] Gleser, L.J. (1992). The Importance of Assessing Measurement Reliability in Multivariate Regression. Journal of the American Statistical Association 87, 696–707.

[4] Hsiao, C. (1983). Identification. In Z. Griliches and M.D. Intriligator eds. Handbook of Econometrics, Vol. I, North-Holland, Amsterdam.

[5] Hsiao, C. (1991). Identification and Estimation of Dichotomous Latent Variables Models Using Panel Data. Review of Economics Studies 58, 717–731.

[6] Wang, L. (1993). Identification and Estimation of Errors-in-Variables Tobit Models. WWZ-Discussion Papers, No. 9304, University of Basel, Switzerland.

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