Economics Letters 38 (1992) 269–273
North-Holland

A nested Tobit analysis for a sequentially censored regression model *

Byung-Joo Lee
University of Colorado, Boulder CO, USA

Received 15 November 1991
Accepted 17 January 1992

This paper proposes a nested Tobit analysis for a sequentially censored regression model. This model is useful for the design and analysis of multiple-step contingent valuation surveys. A consistent maximum likelihood estimation is proposed, and the proper log-likelihood function is derived.

1. Introduction

Limited dependent variable analysis has made rigorous econometric applications possible for a wide variety of survey designs. Survey data often consist in part of qualitative binary choice variables. In contingent valuation surveys, a subset of respondents will also have quantitative values. In such cases, various versions of Tobit model may be appropriate.

There has been extensive research on most of those limited dependent variable models. McFadden (1973) and Amemiya (1973) led the early pioneering work, followed by the more elaborate simultaneous equation models of Nelson and Olson (1978), Heckman (1974), Amemiya (1974) and Lee (1978). Recent research interests shifted toward the semi-parametric or the non-parametric analysis, e.g., by Manski (1975), Powell (1986), Stoker (1986) and Ichimura (1987).

However, for a certain type of survey structure, the existing models are not readily applicable. For example, consider a situation in which a policy maker is interested in providing several different levels of benefit with an additional tax payers' burden. An appropriate survey would then consist of two parts: first, whether or not the respondent is willing to accept the benefit with an additional tax payment; and secondly, if accepting the proposal, how much they would be willing to pay more to get the benefit. For those who responded 'yes', the same questions are repeated for the next level of benefit. For this type of data, the existing models do not properly utilize the

Correspondence to: Byung-Joo Lee, Department of Economics, University of Colorado, Boulder, CO 80309-0256, USA.
* The proposed model was first brought to my attention by my colleague, Charles Howe who used a sequentially censored sample in a contingent valuation survey. I thank him for calling to my attention the importance of this type of model in designing and analyzing surveys.

additional information. A nested logit model is not appropriate because it has a quantitative response. A simple Tobit model is also inappropriate because there are sequentially censored responses from various levels of benefit.

This paper proposes an appropriate estimation model which deals with sequentially censored quantitative responses. The proposed model is called a nested Tobit model which is a hybrid of the Tobit and the nested logit model. The next section develops the nested Tobit model and derives a log-likelihood function. Section 3 provides implications of this model and expands the scope of its applicability.

2. The nested Tobit model

This paper deals with only two-step sequential responses. Even though the multiple-step sequential responses can be modeled theoretically, the computational burden may not warrant its empirical application. [1]

Consider the following two-equation model:

$$\begin{aligned} y_{1i}^* &= x_{1i}'\beta_1 + u_{1i}, \\ y_{2i}^* &= x_{2i}'\beta_2 + u_{2i}, \end{aligned} \tag{1}$$

where $u_i = (u_{1i}, u_{2i})'$ has an i.i.d. bivariate normal distribution, with mean zero and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}.$$

The censored variable y_{1i} is observed if $x_{1i}'\beta_1 + u_{1i} > 0$, and 0 otherwise. Once y_{1i} is observed, y_{2i} is sequentially observed if $x_{2i}'\beta_2 + u_{2i} > 0$, and 0 otherwise. x_{1i} and x_{2i} are $(k_1 \times 1)$ and $(k_2 \times 1)$ observable vectors, and they could be the same set of explanatory variables. This model is similar to Type-2 or Type-3 Tobit models of Amemiya (1984). However, neither of those Tobit models handle the sequential nature of the nested Tobit model properly. The proposed model is summarized as follows:

$$\begin{aligned} y_{1i}^* &= y_{1i} \quad \text{if} \quad x_{1i}'\beta_1 + u_{1i} > 0, \\ & \qquad\qquad \begin{cases} y_{2i}^* = y_{2i} & \text{if} \quad x_{2i}'\beta_2 + u_{2i} > 0 \\ \phantom{y_{2i}^*} = 0 & \text{if} \quad x_{2i}'\beta_2 + u_{2i} \le 0 \end{cases}, \\ &= 0 \quad \text{if} \quad x_{1i}'\beta_1 + u_{1i} \le 0. \end{aligned} \tag{2}$$

[1] Since we are dealing with the normal density function, a multivariate integral evaluation is needed for each step of the multiple sequential responses.

This model can be estimated consistently by maximum likelihood estimation. Define two index variables $d_{1i} = 1(x_{1i}'\beta_1 + u_{1i} > 0)$ and $d_{2i} = 1(x_{2i}'\beta_2 + u_{2i} > 0)$, where the indicator function $1(\lambda)$ is one if λ is true, and zero otherwise. Using these index variables, the likelihood function can be written as follows:

$$L = \prod_{d_{1i}=0} Pr(y_{1i}^* \leq 0) \prod_{d_{1i}=1} \left[\prod_{d_{2i}=0} Pr(y_{2i}^* \leq 0,\, y_{1i}) \prod_{d_{2i}=1} f(y_{2i},\, y_{1i}) \right], \tag{3}$$

where $\prod$ stands for the product over d_{1i} and d_{2i}. This likelihood function is a combination of the probability function and the density function. For a specific form of the likelihood function, consider each term of eq. (3) separately. Since u_i has a bivariate normal distribution, the first term can be written as

$$Pr(y_{1i}^* \leq 0) = Pr(x_{1i}'\beta_1 + u_{1i} \leq 0) = 1 - \Phi\left(\frac{x_{1i}'\beta_1}{\sigma_1}\right), \tag{4}$$

where $\Phi(\cdot)$ is the cumulative normal density function. The second term is

$$Pr(y_{2i}^* \leq 0,\, y_{1i}) = \int_{-\infty}^{0} f(y_{2i}^*,\, y_{1i})\, \mathrm{d}y_{2i}^*. \tag{5}$$

The joint normal density function $f(y_{2i}^*, y_{1i})$ can be decomposed into the product of the conditional normal density function and the marginal normal density function. Then, eq. (5) can be expressed as

$$Pr(y_{2i}^* \leq 0,\, y_{1i}) = f(y_{1i}) \int_{-\infty}^{0} f(y_{2i}^* \mid y_{1i})\, \mathrm{d}y_{2i}^*. \tag{6}$$

In eq. (6), $f(y_{1i})$ is the univariate normal density function with mean $x_{1i}'\beta_1$ and variance σ_1^2, and $f(y_{2i}^* \mid y_{1i})$ is the (conditional) univariate normal density given $y_{1i}^* = y_{1i}$, with mean $x_{2i}'\beta_2 + \sigma_{12}\sigma_1^{-2}(y_{1i} - x_{1i}'\beta_1)$ and variance $(\sigma_2^2 - \sigma_{12}^2\sigma_1^{-2})$. Therefore, the second term of eq. (3) is

$$Pr(y_{2i}^* \leq 0,\, y_{1i}) = \frac{1}{\sigma_1}\phi\left(\frac{y_{1i} - x_{1i}'\beta_1}{\sigma_1}\right)\left[1 - \Phi\left(\frac{x_{2i}'\beta_2 + \sigma_{12}\sigma_1^{-2}(y_{1i} - x_{1i}'\beta_1)}{\left(\sigma_2^2 - \sigma_{12}^2\sigma_1^{-2}\right)^{1/2}}\right)\right], \tag{7}$$

where $\phi(\cdot)$ is the normal density function. The last term of eq. (3) is a bivariate normal density function for y_{2i} and y_{1i} such that

$$f(y_{2i},\, y_{1i}) = \frac{1}{2\pi}|\Sigma|^{-1/2} \exp\left[-\tfrac{1}{2}(y_i - x_i'\beta)'\Sigma^{-1}(y_i - x_i'\beta)\right], \tag{8}$$

where

$$y_i = \begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix}, \quad x_i = \begin{pmatrix} x_{1i} & 0 \\ 0 & x_{2i} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix},$$

and exp(·) is the exponential function. Collecting all these results, we have the following log-likelihood function to estimate β and Σ for the nested Tobit model:

$$\begin{aligned}\mathscr{L}(\beta, \Sigma) = \sum_{i=1}^{n} \Bigg\{ &(1-d_{1i}) \log\left[1 - \Phi\left(\frac{x_{1i}'\beta_1}{\sigma_1}\right)\right] \\ &+ d_{1i}\Bigg[(1-d_{2i}) \log\Bigg(\frac{1}{\sigma_1}\phi\left(\frac{y_{1i} - x_{1i}'\beta_1}{\sigma_1}\right) \\ &\qquad \times \left(1 - \Phi\left(\frac{x_{2i}'\beta_2 + \sigma_{12}\sigma_1^{-2}(y_{1i} - x_{1i}'\beta_1)}{\left(\sigma_2^2 - \sigma_{12}^2\sigma_1^{-2}\right)^{1/2}}\right)\right)\Bigg) \\ &+ d_{2i} \log\left(\frac{1}{2\pi}|\Sigma|^{-1/2} \exp\left[-\tfrac{1}{2}(y_i - x_i'\beta)'\Sigma^{-1}(y_i - x_i'\beta)\right]\right)\Bigg]\Bigg\} \end{aligned} \tag{9}$$

From this log-likelihood function, each parameter of β and Σ is identified separately. These parameters are estimated iteratively using Newton–Raphson method or method of scoring. [2] This log-likelihood function has a univariate normal cumulative density function, which involves a single integral calculation. However, for the multiple-step response model, the log-likelihood function contains a multivariate integral evaluation. This can be quite costly even with high-speed computers.

3. Implications and conclusion

This model is particularly important in a situation where a policy maker would like to know the marginal response rate for different levels of benefit. If we consider $y_i^* = (y_{1i}^*, y_{2i}^*)'$ as indirect utility functions with the relevant socio-economic variables x_i, then a change from y_{1i}^* to y_{2i}^* is a marginal indirect utility, and β can be interpreted as a marginal response rate of the socio-economic variables to the different levels of benefit.

The same model can also be used to determine the optimal fee structure for the different levels of service for different groups. For various options of service, differential service fees can be assigned for different socio-economic groups using the estimated β.

This paper proposes a nested Tobit model for sequentially censored samples. A nested Tobit model is similar to the nested logit model for the sequential choice problem. However, when we have sequential quantitative responses, the existing models do not properly utilize this additional information. A typical approach for the sequentially censored response model is to pool the quantitative responses as if each response were an independent observation. This procedure

[2] I have attempted to derive the analytic expression for the information matrix of eq. (9). However, it is unnecessarily complicated because of the conditional normal density function. Computer evaluates the second derivatives of the log-likelihood function numerically.

incorrectly assumes that each response is independent, and it produces inconsistent parameter estimates. The proposed maximum likelihood function estimates the model parameters consistently, considering the sequential nature of responses.

References

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 variables are truncated normal, Econometrica 42, 999–1012.

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

Heckman, J., 1974, Shadow prices, market wages, and labor supply, Econometrica 42, 679–694.

Ichimura, H., 1987, Estimation of single index models. Ph.D. thesis (MIT, Cambridge, MA).

Lee, L.F., 1978, Unionism and wage rates: A simultaneous equation model with qualitative and limited dependent variables, International Economic Review 19, 415–433.

McFadden, D., 1973, Conditional logit analysis of qualitative choice behaviour, in: P. Zarembka, ed., Frontiers in econometrics (Academic Press, New York).

Manski, C., 1975, Maximum score estimation of the stochastic utility model of choice, Journal of Econometrics 3, 205–228.

Nelson, F.D. and L. Olson, 1978, Specification and estimation of a simultaneous equation model with limited dependent variables, International Economic Review 19, 695–709.

Powell, J., 1986, Symmetrically trimmed least squares estimation for Tobit models, Econometrica 54, 1435–1460.

Stoker, T., 1986, Consistent estimation of scaled coefficients, Econometrica 54, 1461–1481.