

## The Tobit model with a non-zero threshold

RICHARD T. CARSON AND YIXIAO SUN

*Department of Economics, University of California, San Diego*

E-mail: rcarson@ucsd.edu, yisun@ucsd.edu

First version received: August 2006; final version accepted: April 2007

**Summary** The standard Tobit maximum likelihood estimator under zero censoring threshold produces inconsistent parameter estimates, when the constant censoring threshold  $\gamma$  is non-zero and unknown. Unfortunately, the recording of a zero rather than the actual censoring threshold value is typical of economic data. Non-trivial minimum purchase prices for most goods, fixed cost for doing business or trading, social customs such as those involving charitable donations, and informal administrative recording practices represent common examples of non-zero constant censoring threshold where the constant threshold is not readily available to the econometrician. Monte Carlo results show that this bias can be extremely large in practice. A new estimator is proposed to estimate the unknown censoring threshold. It is shown that the estimator is superconsistent and follows an exponential distribution in large samples. Due to the superconsistency, the asymptotic distribution of the maximum likelihood estimator of other parameters is not affected by the estimation uncertainty of the censoring threshold.

**Key words:** *Exponential distribution, Maximum likelihood, Order statistic, Threshold determination.*

### 1. INTRODUCTION

In this paper, we consider the standard Tobit model (Tobin 1958):

$$y_i^* = \alpha + X_i\beta + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (1)$$

where  $y_i^*$  is a latent response variable,  $X_i$  is an observed  $1 \times k$  vector of explanatory variables, and  $\varepsilon_i \sim$  i.i.d.  $N(0, \sigma^2)$  and is independent of  $X_i$ . Instead of observing  $y_i^*$ , we observe  $y_i$ :

$$y_i = \begin{cases} y_i^*, & \text{if } y_i^* > \gamma \\ 0, & \text{if } y_i^* \leq \gamma, \end{cases} \quad (2)$$

where  $\gamma$  is a non-stochastic constant. In other words, the value of  $y_i^*$  is missing when it is less than or equal to  $\gamma$ .

The problem with the standard Tobit model is that  $\gamma$  is often not observed in economic data and is often assumed to be zero in empirical applications. A classic example is willingness to pay for an automobile where  $y_i$  is the observed expenditure on an automobile, which is typically recorded as zero if no automobile is purchased. The unknown censoring threshold,  $\gamma$ , is, in fact, not zero but the lowest available price for a car. Another example is the dollar value of trade

between two countries. It makes little sense to assume that the censoring threshold is zero as it seems highly unlikely that the trade is of a minuscule amount, say 100 or 200 dollars. In this case, the true censoring point depends on the fixed cost of the trade among other factors and may be well above \$0. In both examples, the censoring threshold is not zero. In fact, censoring at a non-zero point is hardly far-fetched, for there is nothing inevitable or magical about zero. It will turn out that the censoring threshold  $\gamma$  is a valuable piece of information.

It is easy to show that the standard maximum likelihood Tobit estimator is inconsistent if the censoring threshold,  $\gamma$ , is not zero but is coded as zero. See, for example, Carson (1988, 1989) and Zuehlke (2003). To consistently estimate the model parameters, we first write  $y_i - \gamma = \max(0, \alpha - \gamma + X_i\beta + \varepsilon_i)$ , a Tobit model with zero censoring point. We then run a standard Tobit on  $y_i - \gamma$  and adjust the estimated intercept by  $\gamma$  to get an estimate of  $\alpha$ . This solution is simple but not useful in empirical work when  $\gamma$  is not known.

In this paper, we consider estimating the unknown censoring threshold by the minimum of the uncensored  $y_i$ 's. We show that the estimator  $\hat{\gamma}$  of  $\gamma$  is superconsistent and asymptotically exponentially distributed. A statistical test is introduced to test whether the threshold is significantly different from a certain value. In addition, it is shown that the maximum likelihood estimator for  $\alpha$ ,  $\beta$  and  $\sigma$  based  $\hat{\gamma}$  is as efficient as the maximum likelihood estimator when the true value of  $\gamma$  is known. Simulation results suggest that the effects of a non-zero  $\gamma$  may be more serious in practice than other previously identified problems with the Tobit model.

The papers that are most closely related to this paper are Carson (1988, 1989) and Zuehlke (2003). Carson (1988, 1989) shows the inconsistency of the maximum likelihood estimator when the censoring point is incorrectly coded and points out that the two-step estimator of Heckman (1976) is consistent. Carson (1988, 1989) also suggests estimating the unknown censoring threshold by the minimum of the uncensored  $y_i$ 's. In a recent paper, Zuehlke (2003) rediscovers these unpublished results and demonstrates via simulations that the asymptotic distribution of the maximum likelihood estimator does not seem to be affected by the estimation of the censoring threshold. One of the main contributions of this paper is to provide a rigorous proof of the latter result.

The rest of the paper is organized as follows. Section 2 describes the estimation procedure and establishes the asymptotic results. The next section provides some simulation evidence. Section 4 contains an empirical application and the final section concludes.

## 2. CONSISTENT AND ASYMPTOTICALLY EFFICIENT ESTIMATOR

To simplify the exposition, we assume that the observations are sorted so that the first  $n_0$  are censored and the last  $n_1 = n - n_0$  are not censored. The likelihood function for the model implied by equations (1) and (2) is usually written (e.g. Cameron and Trivedi 2005) as:

$$L(\alpha, \beta, \sigma) = \prod_{i=1}^{n_0} \Phi\left(\frac{\gamma - \alpha - X_i\beta}{\sigma}\right) \prod_{i=n_0+1}^n \frac{1}{\sigma} \phi\left(\frac{y_i - \alpha - X_i\beta}{\sigma}\right), \tag{3}$$

where  $\Phi(\cdot)$  and  $\phi(\cdot)$  are the standard normal cumulative distribution function and probability density function (pdf), respectively.

Let  $X = (X'_{(0)}, X'_{(1)})'$ ,  $y = (y'_{(0)}, y'_{(1)})'$ , where the subscripts on  $X$  and  $y$  indicate whether they are from the censored,  $n_0$ , or uncensored,  $n_1$ , portion of the sample. To obtain the maximum likelihood estimators of  $\alpha$ ,  $\beta$  and  $\sigma$ , we follow Carson (1988, 1989) and Zuehlke (2003) and first

estimate  $\gamma$  by  $\hat{\gamma}$ :

$$\hat{\gamma} = \min \{y_{(1)}\}. \tag{4}$$

We claim that  $\hat{\gamma}$  is superconsistent, converging to the true value of  $\gamma$  at the rate of  $1/n$ . To see this, note that conditioning on  $X_{(1)}$ , the elements of  $y_{(1)}$  follow truncated distributions with

$$P(y_{(1),i} < u | X_{(1),i}) = \frac{\Phi[(u - \alpha - X_{(1),i}\beta)/\sigma] - \Phi[(\gamma - \alpha - X_{(1),i}\beta)/\sigma]}{1 - \Phi[(\gamma - \alpha - X_{(1),i}\beta)/\sigma]}. \tag{5}$$

Conditioning on  $n_1$  and  $X$ , we have, for any  $\delta > 0$ ,

$$\begin{aligned} &P(\min \{y_{(1)}\} > \gamma + \delta/n | X, n_1) \\ &= P(y_{(1),1} > \gamma + \delta/n, y_{(1),2} > \gamma + \delta/n, \dots, y_{(1),n_1} > \gamma + \delta/n | X, n_1) \\ &= \prod_{i=1}^{n_1} \left\{ \left( 1 - \frac{\Phi[(\gamma + \delta/n - \alpha - X_{(1),i}\beta)/\sigma] - \Phi[(\gamma - \alpha - X_{(1),i}\beta)/\sigma]}{1 - \Phi[(\gamma - \alpha - X_{(1),i}\beta)/\sigma]} \right) \right\} \\ &= \prod_{i=1}^{n_1} \left\{ 1 - \frac{\delta}{n\sigma} \frac{\phi[(\gamma - \alpha - X_{(1),i}\beta)/\sigma] + o_p(1)}{1 - \Phi[(\gamma - \alpha - X_{(1),i}\beta)/\sigma]} \right\} \end{aligned} \tag{6}$$

as  $n \rightarrow \infty$ , where the  $o_p(1)$  term holds uniformly over the support of  $X_i$  as  $\phi(\cdot)$  is bounded.

Let

$$\mu_i = \frac{1/\sigma \phi[(\gamma - \alpha - X_{(1),i}\beta)/\sigma]}{(1 - \Phi[(\gamma - \alpha - X_{(1),i}\beta)/\sigma])} \text{ and } \mu = E\mu_i, \tag{7}$$

where the expectation is taken with respect to  $X_{(1),i}$ . If  $E\|X_{(1),i}\|^{2+\Delta} < \infty$  for some small  $\Delta > 0$ , then  $E\|\mu_i\|^{2+\Delta} < \infty$ . This is because for any  $z \in \mathbb{R}$ , there exists a constant  $C > 0$  independent of  $z$  such that  $|\phi(z)/(1 - \Phi(z))| \leq C|z|$ . The moment condition on  $\mu_i$  helps us to bound  $\max|\mu_i|$ : for any  $M > 0$ ,

$$P(\max |\mu_i| > n^{1/(2+\Delta)}M) \leq nP(|\mu_i| > n^{1/(2+\Delta)}M) = (E\|\mu_i\|^{2+\Delta})M^{-2-\Delta}. \tag{8}$$

Hence  $\max|\mu_i| = O_p(n^{1/(2+\Delta)})$ .<sup>1</sup> Now

$$\begin{aligned} &\frac{1}{n_1} \log P(\min \{y_{(1)}\} > \gamma + \delta/n | X, n_1) \\ &= \frac{1}{n_1} \sum_{i=1}^{n_1} \log \left[ 1 - \frac{\delta}{n} \mu_i + o_p\left(\frac{1}{n}\right) \right] \\ &= \frac{1}{n_1} \sum_{i=1}^{n_1} \frac{\delta}{n} \mu_i + o_p\left(\frac{1}{n}\right) + O\left(\frac{\max |\mu_i|^2}{n^2}\right) \\ &= \frac{1}{n_1} \sum_{i=1}^{n_1} \frac{\delta}{n} \mu_i + o_p\left(\frac{1}{n}\right) = \frac{\mu\delta}{n}(1 + o_p(1)), \end{aligned} \tag{9}$$

<sup>1</sup> We derive this result using a crude maximal inequality under the assumption that  $E\|X_{(1),i}\|^{2+\Delta} < \infty$ . The moment condition may be relaxed if we use a refined maximal inequality.

where the last equality holds by a law of large numbers. Then conditioning only on  $n_1$ ,

$$P(\min\{y_{(1)}\} > \gamma + \delta/n|n_1) = \exp\left[\frac{n_1}{n}\mu\delta(1 + o_p(1))\right], \tag{10}$$

as  $n \rightarrow \infty$ . Note that as  $n \rightarrow \infty$ ,

$$\frac{n_1}{n} = \frac{\sum_{i=1}^n \mathbf{1}\{y_i^* > \gamma\}}{n} \rightarrow E\{1 - \Phi[(\gamma - \alpha - X_i\beta)/\sigma]\} \tag{11}$$

by a law of large numbers. As a consequence,

$$P(\min\{y_{(1)}\} > \gamma + \delta/n|n_1) \rightarrow \exp(-\eta\delta), \tag{12}$$

where

$$\eta = E\{1 - \Phi[(\gamma - \alpha - X_i\beta)/\sigma]\} E\frac{1/\sigma\phi[(\gamma - \alpha - X_{(1),i}\beta)/\sigma]}{1 - \Phi[(\gamma - \alpha - X_{(1),i}\beta)/\sigma]}. \tag{13}$$

Since the right-hand side of (12) does not depend on  $n_1$ , it is also the limit of the unconditional probability. Therefore, we have shown that

$$\lim_{n \rightarrow \infty} P(n(\min\{y_{(1)}\} - \gamma) < \delta) = 1 - \exp(-\eta\delta) \text{ for } \delta > 0, \tag{14}$$

an exponential distribution with parameter  $\eta$ .

The mean of this distribution is  $1/\eta$  and the variance is  $1/\eta^2$ . As  $\eta$  increases, the limiting distribution becomes more concentrated around the origin. Heuristically, the first factor in  $\eta$  measures the probability of not censoring while the second factor measures how likely an observation is close to  $\gamma$  given that it is not censored. Combining these two factors, we can regard  $\eta$  as a measure of the likelihood of getting an observation that is close to  $\gamma$ . Therefore, the larger  $\eta$  is, the closer  $\min\{y_{(1)}\}$  is to  $\gamma$  and the more concentrated the limiting distribution is.

Note that while we have used the independence of  $\{y_i\}$  and the differentiability of the normal CDF to derive the limiting distribution, we have not assumed that the  $y_i$ 's are identically distributed as do most of the results on asymptotics of order statistics.

Since the first part of the likelihood function in (3) is increasing in  $\gamma$ , where the second part is independent of  $\gamma$ ,  $\hat{\gamma}$  is the maximum likelihood estimator of  $\gamma$  under the constraint that  $\gamma \leq \min\{y_{(1)}\}$ . The jointly sufficient statistics for (3) are  $\hat{\gamma}, \alpha_{\hat{\gamma}}, \beta_{\hat{\gamma}}, \sigma_{\hat{\gamma}}$  where  $[\alpha_{\hat{\gamma}}, \beta_{\hat{\gamma}}, \sigma_{\hat{\gamma}}]$  are the maximum likelihood estimates assuming  $\gamma = \hat{\gamma}$ . In view of the superconsistency of  $\hat{\gamma}$ , a standard textbook argument shows that  $\alpha_{\hat{\gamma}}, \beta_{\hat{\gamma}}, \sigma_{\hat{\gamma}}$  are asymptotically normal with the asymptotic variance identical to that for  $[\alpha_{\gamma}, \beta_{\gamma}, \sigma_{\gamma}]$  when  $\gamma$  is known. The superconsistency of  $\hat{\gamma}$  is a well-known consequence of the dependence of the support of  $y$  on  $\gamma$  (e.g. Schervish 1995, pp. 408).

Computationally the proposed estimator is simple: retain the  $\min\{y_{(1)}\}$  and estimate the Tobit model by maximum likelihood assuming  $\gamma = \min\{y_{(1)}\}$ . Report the usual asymptotic covariance matrix for  $[\alpha_{\hat{\gamma}}, \beta_{\hat{\gamma}}, \sigma_{\hat{\gamma}}]$ .

What we have done is equivalent to respecifying (3) to explicitly include the restriction on the support of  $y$  so that the likelihood function becomes:

$$L(\alpha, \beta, \sigma) = \prod_{i=1}^{n_0} \Phi\left(\frac{\gamma - \alpha - X_i\beta}{\sigma}\right) \prod_{i=n_0+1}^n \frac{1}{\sigma} \phi\left(\frac{y_i - \alpha - X_i\beta}{\sigma}\right) \mathbf{1}\{y_i > \gamma\}. \tag{15}$$

The above likelihood function is also given in Carson (1988, 1989) and Zuehlke (2003). The important difference between this likelihood function and that given in (3) is that the likelihood



### 3. MONTE CARLO EXPERIMENTS

In this section, we first investigate the magnitude of the finite sample bias of the maximum likelihood estimator under the standard Tobit specification when  $\gamma \neq 0$ , and then we investigate the accuracy of the asymptotically exponential approximation for the distribution of  $\hat{\gamma}$ . Finally, we examine the size and power properties of the  $T$  test.

Inspired by the automobile purchase problem, we let

$$y_i^* = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i, \quad (21)$$

where  $\alpha = 6000$ ,  $\beta_1 = \beta_2 = 1$ ,  $x_{1i}$  and  $x_{2i}$  are normal random variables each with a mean 5000 and a standard deviation of 500. The correlation between  $x_{1i}$  and  $x_{2i}$  is set at 0.2. The error terms,  $\varepsilon_i$ 's, are generated to be normally distributed with mean 0 and standard deviation 1000. Thus the expected value of  $y_i^*$  is 16 000. The observations are censored at  $\gamma = 15\,000$  which produces samples with roughly 20% of the 500 observations being censored.

Table 1 presents the average value and standard error of each estimator over 5000 replications. OLS estimates are given for the actual  $y^*$ , only the uncensored observations  $y_{(1)}$ , and the incorrectly coded  $y^o$  which has zero recorded for the censored values. Tobit estimates are given using  $\gamma$ 's of 0, 2500, 5000, 7500, 10 000, 12 500, and 15 000, the true value. Estimates are also given for two consistent techniques, the Tobit using  $\gamma = \min\{y_{(1)}\}$  and Heckman's two-step procedure (Heckman 1976). The bias in the constant term estimated using the standard Tobit model (i.e.  $\gamma = 0$ ) is over 550%; and the bias in the estimates of the slope coefficients is over 400%. For the design matrix and parameter values used, the bias in each of the Tobit coefficients declines at very close to a linear rate until the  $\gamma$  assumed takes on its true value of 15 000 at which point the bias is essentially zero. A linear regression of the bias in  $\hat{\alpha}$  on 15 000 minus the assumed  $\gamma$  yields  $-3.0904 \times (15\,000 - \gamma)$  with  $R^2 = 0.999$ . For the biases in  $\beta_1$  and  $\beta_2$  the regression equation is  $0.000283 \times (15\,000 - \gamma)$  with  $R^2 = 0.999$ . It is important to note that the large bias in the parameters exhibited by this Monte Carlo simulation is not strictly due to  $\gamma = 15\,000$ . Since the bias is directly related to  $|\gamma/\sigma|$ , the identical percentage bias could have been generated using say  $\gamma = 0.15$ ,  $\sigma = .01$  and an appropriate rescaling of the  $X$  matrix.

The OLS estimates using  $y^o$ , in which the censored values are coded zero, are severely biased although less so than the standard Tobit model with  $\gamma$  set equal to 0, and appear similar to a Tobit model with an assumed  $\gamma$  between 2500 and 5000. The OLS estimates using only  $y_{(1)}$ , the uncensored observations, are even less biased but the pattern of parameter estimates is very different with the intercept biased upward and the slope coefficients biased downward.

The Tobit model using  $\gamma = \min\{y_{(1)}\}$  produces consistent parameter estimates which for all practical purposes are indistinguishable from the Tobit model using  $\gamma = 15\,000$ , the true value. The Heckman estimator is also consistent. For the two slope coefficients its loss in efficiency is large, being more than 100% for both coefficients. The Heckman estimator's loss in efficiency is even larger for the estimates of  $\alpha$  and  $\sigma$ , being 208% and more than 1000%, respectively.

Next, we investigate the accuracy of the asymptotically exponential approximation. Using the same data generating process as before, we simulate the statistic  $T$  defined in (17). We employ the kernel smoothing method to estimate the probability density of  $T$ . The underlying bandwidths are based on the plug-in rule proposed by Sheather and Jones (1991). For values of  $T$  at the boundary of its support, we use the data reflection method of Hall and Wehrly (1991) to correct for the boundary bias. Figure 1 presents the kernel density estimate (KDE) based on 5000 replications with sample size 500. In the same figure, we also graph the pdf of the standard exponential

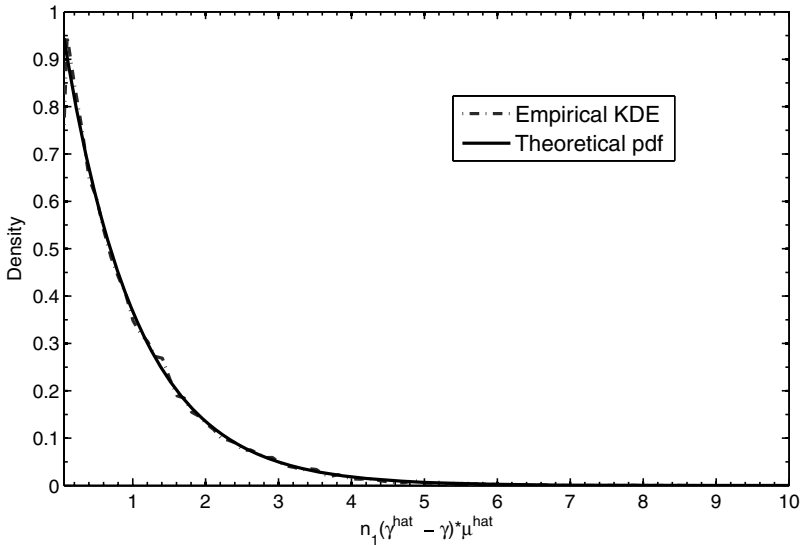
**Table 1.** Finite sample performances of different estimators.

Method	$\alpha$	$\beta_1$	$\beta_2$	$\sigma$
TRUE	6000	1.0000	1.000	1000
OLS ( $y^*$ )	5984.7 (581.2)	1.0003 (0.0902)	1.0028 (0.0912)	997.52 (32.01)
OLS ( $y_{(1)}$ )	9569.8 (596.0)	0.6761 (0.0892)	0.6803 (0.0891)	820.52 (29.75)
OLS ( $y^o$ )	-29 653 (3484.1)	4.2583 (0.5325)	4.2593 (0.5305)	5948.70 (186.38)
ML ( $\gamma = 0$ )	-40 742 (5235.7)	5.2859 (0.7236)	5.2875 (0.7243)	7420.00 (359.86)
ML ( $\gamma = 2500$ )	-32 764 (4418.6)	4.5539 (0.6116)	4.5558 (0.6123)	6280.40 (304.66)
ML ( $\gamma = 5000$ )	-24 797 (3604.4)	3.8229 (0.5000)	3.8250 (0.5008)	5145.20 (249.39)
ML ( $\gamma = 7500$ )	-16 850 (2795.5)	3.0938 (0.3892)	3.0961 (0.3901)	4018.5 (194.04)
ML ( $\gamma = 10\,000$ )	-8946.1 (1998.5)	2.3687 (0.2803)	2.3713 (0.2813)	2909.70 (138.56)
ML ( $\gamma = 12\,500$ )	-1169.8 (1236.4)	1.6554 (0.1767)	1.6583 (0.1777)	1850.80 (83.31)
ML ( $\gamma = 15\,000$ )	5975.9 (642.1)	1.0008 (0.0962)	1.004 (0.0968)	997.69 (37.46)
ML $\gamma = \min(y_{(1)})$	5996.5 (641.4)	0.9989 (0.0960)	1.0021 (0.0966)	995.72 (37.35)
Heckit	6078.6 (1944.7)	0.9908 (0.1930)	0.9947 (0.1940)	987.22 (494.06)

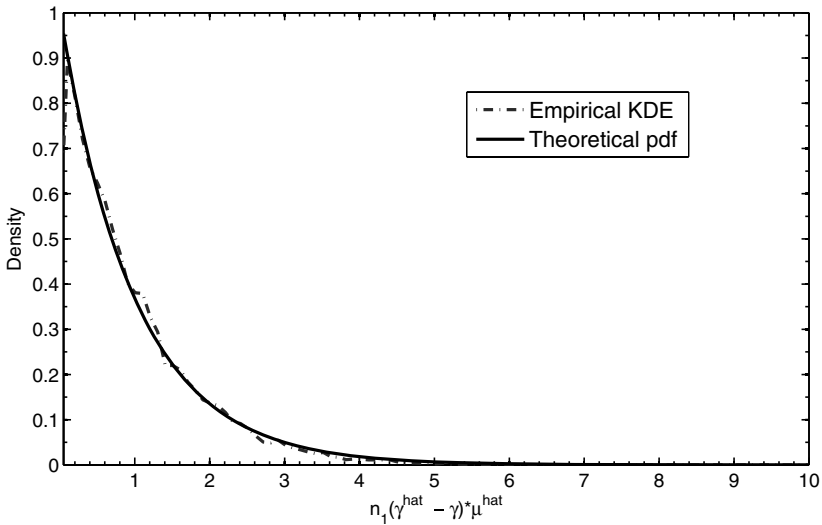
*Note.* Standard errors are given in parentheses. ML ( $\gamma = z$ ) is the Tobit maximum likelihood estimator obtained with  $\gamma$  assumed to be  $z$ . The true value for  $\gamma$  is 15 000. Results are based on 5000 replications with a sample size of 500.

distribution. It is clear that the empirical KDE is close to the theoretical pdf, suggesting that the asymptotic distribution provides a good approximation to the finite sample distribution. Figure 2 is the same as Figure 1 except that the same size is 100. Again the approximation is accurate even for this smaller sample size. It should be pointed out that in both figures the empirical KDE is smaller than the theoretical pdf near the boundary of the support. This is partly due to the remaining boundary bias of the kernel density estimator.

Finally, we examine the finite sample properties of the  $T$  test. The data generating process is the same as before except that several values of the intercept are considered, that is,  $\alpha = 4000$ ,



**Figure 1.** Kernel density estimate of  $T = n_1(\hat{\gamma} - \gamma)\hat{\mu}$  with sample size 500.



**Figure 2.** Kernel density estimate of  $T = n_1(\hat{\gamma} - \gamma)\hat{\mu}$  with sample size 100.

5000, 6000, 7000 and 8000. All else being equal,  $\mu$  decreases with  $\alpha$ . Therefore, as  $\alpha$  increases, it is less likely to get many uncensored observations that are close to the censoring point. Table 2 presents the size and power of the  $T$  test. The column with  $\gamma = 15000$  contains the size of the test. The empirical size of the  $T$  test is close to the nominal size 5%. This result confirms the above simulation evidence that the asymptotic distribution provides a good approximation to the finite sample distribution of the  $T$  statistic. Other columns of Table 2 contain the power of the test. The simulation results show that the power decreases as  $\alpha$  increases. This is not surprising



**Table 2.** Size and power of the  $T$  test at the 5% level with  $H_0: \gamma = 15\,000$  and  $H_1: \gamma > 15\,000$ .

	$\gamma = 15\,000$	$\gamma = 15\,025$	$\gamma = 15\,050$	$\gamma = 15\,075$	$\gamma = 15\,100$
$\alpha = 4000$	0.0510	0.8355	1.0000	1.0000	1.0000
$\alpha = 5000$	0.0530	1.0000	1.0000	1.0000	1.0000
$\alpha = 6000$	0.0507	0.9167	1.0000	1.0000	1.0000
$\alpha = 7000$	0.0517	0.1733	0.5794	0.9985	1.0000
$\alpha = 8000$	0.0794	0.1022	0.1347	0.1756	0.2338

Note. Results are based on 10 000 replications with a sample size of 500.

because when  $\alpha$  increases, the censoring point moves to the lower tail of the distribution of  $y_i^*$ . As an example, when  $\alpha = 8000$  and  $\gamma = 15\,025$ , only 0.9% of the observations are censored. This being said, the  $T$  test is very powerful when the censoring point is reasonably away from the left tail of  $y_i^*$ .

#### 4. EMPIRICAL APPLICATION: DONATION TO CHARITY

The simulation exercise in the previous section is a stylized version of the durable goods case for which the Tobit model was originally developed and is still in common usage. The new TSP 5.0 manual (Hall and Cummins 2005) for instance features a simple automobile example with censoring at zero. In that case  $\gamma$  and, more particularly,  $|\gamma/\sigma|$  is likely to be relatively large and hence represent the ideal case for the method we propose to be useful. In this section, we look at an empirical example from another common application of the Tobit model, charitable contributions (e.g. Reece 1979). In this case one might well expect *a priori* that the minimum observed positive contribution in a data set is fairly small and that there is a fairly wide range of contributions suggesting that  $\sigma$  is large. As a result, the Tobit model with a zero threshold does not seem to be seriously misspecified. However, we show that the Tobit model with non-zero threshold has a larger likelihood function and better out-of-sample forecasting performance than the Tobit model with zero threshold.

The dataset we used comes from Franses and Raap (2001, section 2.2.5). It consists of observations of 4268 individuals concerning the response to a charity direct mail appeal. The dependent variable is the donation in response to the latest charity mailing. The dummy variable 'Resp' indicates whether an individual responded to the latest charity mailing. The rest of the variables will be used as proxies for the willingness to donate. They are a 0–1 dummy variable indicating whether an individual responded to the previous mailing, the number of weeks since the last response, the average number of mailings received per year, the percentage of responses in the past, the average donation in the past and the amount donated in the last response. These variables are listed in Table 3. Table 4 reports the minimum, average, standard deviation, maximum of all variables for the two subsamples: the subsample who responded to the latest mailing and the subsample who did not. The average donation for those who responded is about 18 Dutch guilders and the standard deviation is 18.96. There are a few individuals who donated more than 200 guilders. Compared with the individuals who responded, individuals who did not respond had lower responding rate and longer responding time, and made a smaller amount of donation in the past.

All individuals in the sample had responded at least once to the previous mailings. This suggests that the willingness to donate is most likely to be positive, as a negative willingness

**Table 3.** Dependent variable and explanatory variables.

Variables	Description
Donation	Donation in Response to the Latest Mailing, Dep. Var.
Resp	Response to the Latest Mailing (Dummy Variable)
Resp_Previous_Mail	Response to the Previous Mailing (Dummy)
Weeks_Last_Resp	Weeks Since Last Response
Percent_Resp	Percentage of Responded Mailings
Mails_Per_Year	Average Number of Mailings per Year
Donation_Last_Resp	Donation in the Last Response
Aver_Past_Donation	Average Donation in the Past

**Table 4.** Characteristics of the dependent variable and explanatory variables.

Variables	No Response <sup>a</sup>				Response <sup>b</sup>			
	Minimum	Average	S.D.	Maximum	Minimum	Average	S.D.	Maximum
Donation	0	0	0	0	2	18.613	18.96	250
Resp	0	0.6000	0.4900	1	0	0.4000	0.4900	1
Resp_Previous_Mail	0	0.2073	0.4055	1	0	0.5261	0.4995	1
Weeks_Last_Resp	13.14	72.10	47.41	195	13.14	39.47	30.02	195
Percent_Resp	9.09	39.27	21.46	100	11.11	62.19	24.49	1
Mails_Per_Year	0.25	1.99	0.65	3.50	0.25	2.15	0.68	3.50
Donation_Last_Resp	1	17.04	22.00	750	2	23.81	242.32	10 000
Aver_Past_Donation	1	16.831	20.73	750	1.83	20.36	121.82	5005

<sup>a</sup>Summary statistics for those who did not respond to the latest charity mailing.

<sup>b</sup>Summary statistics for those who responded to the latest charity mailing.

implies that an individual wanted to take some money from the charity fund instead of making a contribution. This observation leads to the following model specification:

$$y_i = \begin{cases} y_i^*, & \text{if } y_i^* \geq \gamma^*, \\ 0, & \text{if } y_i^* < \gamma^*, \end{cases} \tag{22}$$

where  $y_i$  is the observed amount of donation and  $y^*$  is the latent variable that measures the willingness to donate. We assume that  $y^*$  satisfies

$$y_i^* = \exp(x_i \beta + \varepsilon_i), \tag{23}$$

where  $\varepsilon_i$  is i.i.d.  $N(0, \sigma^2)$  and independent of  $\{x_i\}_{i=1}^n$ . Here  $x_i$  includes all the proxy variables described above but with the average donation in the past and the amount donated in the last response in logarithm forms.

The model given in (22) and (23) is a log-normal variant of the standard Tobit model. It can be rewritten as

$$y_i = \begin{cases} y_i^*, & \text{if } \log y_i^* \geq \gamma, \\ 0, & \text{if } \log y_i^* < \gamma. \end{cases} \tag{24}$$

**Table 5.** Estimation and forecasting results for the Tobit model with zero and non-zero censoring points.

Explanatory Variables	Non-Zero Censoring		Zero Censoring	
	Parameter	<i>t</i> -stat	Parameter	<i>t</i> -stat
Constant	-2.4514	-10.94	-3.6831	-12.50
Resp_Previous_Mail	0.0950	1.00	0.1309	1.04
Weeks_Last_Resp	-0.0100	-8.00	-0.0134	-8.06
Mails_Per_Year	0.1869	3.49	0.2496	3.53
Percent_Resp	2.8875	15.36	3.8296	15.45
Log(Aver_Donation)	0.6092	11.60	0.6345	9.14
$\sigma$		1.8704		2.4760
Log-likelihood value		-8840.4		-9302.7
LR ratio		1024.01		998.05
MAFE		5.1246		5.1864
RMSFE		21.670		79.509

Note. The sample size is 4000 and the number of censored observations is 2393.

The threshold  $\gamma$ , defined as  $\log(\gamma^*)$ , is unknown and possible non-zero. Allowing for an unknown and non-zero threshold makes the log-normal variant of the Tobit model a more attractive option as zero is unlikely to be a plausible option in this case as it simply corresponds to censoring at 1 in the latent variable’s original units.

A common practice in research on charitable donations is to model  $\log(a + y_i)$  using the Tobit with a zero threshold or a linear regression model. Here the constant  $a$  is chosen by the researchers and is often quite *ad hoc*. For example, the choice of  $a = 10$  has long been common in the literature on charitable donations (e.g. Choe and Jeong 1993). While the *ad hoc* approach works reasonably well for our example, there is no reason to believe that adding an *ad hoc* correction factor to the dependent variable will work well in general.

We proceed to implement our proposed model. We drop  $\log(\text{Aver\_Past\_Donation})$  from the vector  $x_i$  as its correlation with  $\log(\text{Donation\_Last\_Resp})$  is extremely high ( $>0.95$ ). A simple plot of  $y_i$  suggests that the original data set has been systematically organized according to some criteria. To make a fair comparison of the out-of-sample performance of different models, we randomly permute the original data set so that the individuals are not ordered in any way. We use the first 4000 individuals to estimate different models and the remaining 268 individuals to evaluate their out-of-sample forecasting performance.

The log-likelihood function for our models is

$$\sum_{y_i=0} \log \Phi\left(\frac{\gamma - x_i \beta}{\sigma}\right) - \sum_{\log(y_i) > \gamma} \left\{ \frac{1}{2} \log(2\pi \sigma^2) + \frac{(\log(y_i) - x_i \beta)^2}{2\sigma^2} + \log(y_i) \right\}, \quad (25)$$

where  $\gamma = \log(\gamma^*)$ . Table 5 represents the maximum likelihood estimate results of the Tobit models with zero and non-zero thresholds. The non-zero threshold  $\gamma$  is estimated using the minimum of the non-zero  $\log(y_i)$ ’s, that is

$$\hat{\gamma} = \min_{y_i > 0} \{\log y_i\}. \quad (26)$$

It is clear that the Tobit model with non-zero threshold has a larger likelihood function than the Tobit model with zero threshold. To formally test  $H_0 : \gamma = 0$  against  $H_1 : \gamma > 0$ , we compute the  $T$  statistic

$$T = n_1 \hat{\mu} \min_{y_i > 0} \{\log y_i\}, \tag{27}$$

where  $\hat{\mu}$  is defined in equation (18). We find that  $T = 437.56$ . Comparing this with the critical value from the standard exponential distribution, we reject the null hypothesis at the 1% level. Alternatively, we can construct the 99% confidence interval for  $\gamma$ , which is [0.686, 0.693]. The confidence interval does not contain the origin, which is consistent with the hypothesis testing result. In addition, the confidence interval is very narrow, suggesting little loss in not knowing the threshold.

For both models, the estimated coefficients have the expected signs. All coefficients are significant at the 1% level except the dummy variable indicating whether an individual responded to the previous mailing. For each of the models, we perform the likelihood ratio test for the joint significance of all the slope coefficients. The respective LR statistics are 1024.01 and 998.05. Comparing these with the  $\chi^2_5$  critical values, we reject the null that the slope coefficients are jointly zero. The typical summary statistics suggest that either model provides a reasonable summary of the data at hand. However, there is a sizeable difference in the estimated coefficients across the two models. The results of Table 5 show that the Tobit model with a zero threshold overestimates the effect of all the proxy variables. Of particular note is that the coefficient on the key variable in the model Percent.Resp, the percent of previous mailing that the person responded to, is over 30% larger in the model with the zero threshold. From a direct mail perspective, this would cause one to substantially over estimate the value of being able to selectively sampling from a database based on donor propensity to contribute in the past. Interestingly, most of the other coefficients are also overestimated by roughly 30% except for the coefficient on Log(Aver\_Donation), the log of the average donation in the past, which shows little change in the magnitude of the coefficient but a substantial decline in the corresponding  $t$ -statistic from 11.60 to 9.14.

To compare the forecasting performance of the two models, we compute the mean of  $y$  for given  $x$  as follows:

$$\begin{aligned} E(y|x) &= P(x\beta + \varepsilon \geq \gamma) E(\exp(x\beta + \varepsilon) | x\beta + \varepsilon \geq \gamma) \\ &= \left\{ 1 - \Phi\left(\frac{\gamma - x\beta}{\sigma}\right) \right\} \exp(x\beta) E(\exp(\varepsilon) | \varepsilon \geq \gamma - x\beta) \\ &= \left\{ 1 - \Phi\left(\frac{\gamma - x\beta}{\sigma}\right) \right\} \exp(x\beta) \left\{ 1 - \Phi\left(\frac{\gamma - x\beta}{\sigma}\right) \right\}^{-1} \\ &\quad \times \int_{\gamma - x\beta}^{\infty} \exp(\varepsilon) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right) d\varepsilon \\ &= \exp(x\beta) \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\gamma - x\beta}^{\infty} \exp\left(-\frac{(\varepsilon - \sigma^2)^2}{2\sigma^2} + \frac{\sigma^4}{2\sigma^2}\right) d\varepsilon \\ &= \left\{ \exp\left(x\beta + \frac{\sigma^2}{2}\right) \right\} \left\{ 1 - \Phi\left(\frac{\gamma - x\beta - \sigma^2}{\sigma}\right) \right\}. \end{aligned} \tag{28}$$

We forecast  $y$  according to

$$\hat{y} = \left\{ \exp \left( x\hat{\beta} + \frac{\hat{\sigma}^2}{2} \right) \right\} \left\{ 1 - \Phi \left( \frac{\hat{y} - x\hat{\beta} - \frac{\hat{\sigma}^2}{2}}{\hat{\sigma}} \right) \right\}. \quad (29)$$

For the 268 individuals in the hold-out sample, we obtain a Root Mean Squared Forecasting Error (RMSFE) of 21.670 for the Tobit model with non-zero threshold and 79.509 for the Tobit model with zero threshold. Hence, the RMSFE performance of the Tobit model with non-zero threshold is much better than that of the Tobit model with zero threshold.

We have also used the Mean Absolute Forecasting Error (MAFE) to evaluate the forecasting performance. For this criterion we assume the conditional median of  $\varepsilon$  given  $x$  to be zero and use the conditional median of  $y$  as our forecast. Since  $y = \exp(x\beta + \varepsilon) \mathbf{1}\{\exp(x\beta + \varepsilon) \geq \gamma^*\}$  is a non-decreasing function of  $(x\beta + \varepsilon)$ ,

$$\text{median}(y|x) = \exp(x\beta) \mathbf{1}\{x\beta \geq \gamma\} = \begin{cases} \exp(x\beta), & \text{if } x\beta \geq \gamma, \\ 0, & \text{if } x\beta < \gamma. \end{cases} \quad (30)$$

We forecast  $y$  according to  $\hat{y} = \exp(x\hat{\beta}) \mathbf{1}\{x\hat{\beta} \geq \hat{\gamma}\}$ .

The MAFE does not show as large an advantage for the non-zero threshold model although it still dominates the zero threshold model. This is not surprising given that much of the improvement in predictive power is concentrated in the tails of the distribution. To check the robustness of our result, we model  $y_i$  itself rather than its logarithm using the Tobit model. Our qualitative results for the zero threshold and non-zero threshold models given above remain valid.

## 5. DISCUSSION AND CONCLUDING REMARKS

The censoring threshold plays a non-trivial role in the Tobit model. While this has long been recognized if the censoring threshold differs for each agent or is stochastic (Nelson 1977), it has not been recognized in the typical case where the censoring threshold is assumed to be constant and non-stochastic. The special case where  $\gamma = 0$  results in important simplifications in the algebra of the Tobit problem.

In most of the situations where Tobit models have been used in practice it is unlikely that  $\gamma$  is actually equal to zero. Examples include purchases of a wide variety of goods ranging from cigarettes to automobiles, trade flows, charitable giving, and administrative records involve wages and health expenditures. Economics is unique among disciplines using censored and truncated distributions in that little attention has been paid to the threshold point. The extent to which existing Tobit estimates from empirical studies are biased is a function of the extent to which  $|\gamma/\sigma|$  is different from zero. How important this is in practice is, of course, an open question and empirical studies do not typically report the statistics which would be necessary to calculate this term. Our empirical application suggests that the effects of estimating the censoring threshold are more subtle in the case where  $|\gamma/\sigma|$  is small but still potentially important from both an inference and predictive perspective.

Our Monte Carlo results suggest that non-zero censoring thresholds may be a more important source of bias than non-normality or heteroscedasticity (e.g. Wooldridge 2002, pp. 533). If one is not prepared to give up the assumption that the censoring threshold is a constant and explicitly model it in the manner of Nelson (1977), then there is a simple recommendation for applied work:

set  $\gamma$  equal to the minimum of the uncensored dependent variable. In large samples there is no efficiency loss relative to knowing and using the correct  $\gamma$ . In order to ensure the observations that achieve  $\min\{y_{(1)}\}$  are treated as uncensored observations, we may actually need to set  $\gamma$  slightly less than  $\min\{y_{(1)}\}$  when using a canned statistical package. In other words, we set  $\gamma = \min\{y_{(1)}\} - \textit{eps}$  where *eps* is the distance between two closest floating point numbers or a number that is much smaller than  $\min\{y_{(1)}\}$ . Some statistical package such as STATA uses the minimum as the left censoring point when the censoring threshold is not specified. However, the minimum is taken over all the observations. Hence, if censored observations are coded as zeros as in typical economic datasets, the default tobit command in STATA will set  $\gamma = 0$  and may produce incorrect results. Our recommendation is to always specify the censoring point when using a canned statistical package.

It is straightforward to see that our simple approach also applies to truncation regression with an unknown truncating point. Our results can be easily extended to the right censoring or truncation cases. In addition, our simple approach can be used to analyse datasets in which the unknown censoring/truncating points differ according to some exogenously specified variable such as state of residence. In this case, we can estimate the censoring/truncating point for each state by the minimum of the uncensored/untruncated dependent variable within each state. As long as the number of observations for each state approaches infinity as the sample size increases, our proof for the asymptotic exponential distribution goes through without modification. Under the assumption of cross sectional independence, the estimators of the censoring/truncating points are i.i.d. exponential in large samples, which gives us a convenient way to test whether differences in the censoring/truncating points across states are significantly different.

## ACKNOWLEDGMENTS

We would like to thank John Horowitz, Whitney Newey, Franco Peracchi, Dan Steinberg, Shankar Subramanian, Pravin Trivedi, Halbert White, and an anonymous referee for helpful comments. We thank Peter Kennedy for pointing out a reference. Correspondence to Yixiao Sun, Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508.

## REFERENCES

- Cameron, A.C. and R.K. Trivedi (2005). *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.
- Carson, R.T. (1988). The Tobit Model with a Non-zero Censoring Threshold, Working Paper 88-37, Department of Economics, University of California, San Diego.
- Carson, R.T. (1989). The Tobit Model with a Non-zero Censoring Threshold, Paper presented at the winter meeting of the Econometric Society, Atlanta, December, 1989.
- Choe, Y. and J. Jeong (1993). Charitable contributions by low- and middle-income taxpayers: Further evidence with a new method. *National Tax Journal* 46, 33–9.
- Franses, F.H. and R. Raap (2001). *Quantitative Models in Marketing Research*. Cambridge, UK: Cambridge University Press.
- Hall, B. H. and C. Cummins (2005). *TSP 5.0 Reference Manual*. Palo Alto, California: TSP International.
- Hall, P. and T.E. Wehrly (1991). A geometrical method for removing edge effects from kernel-type nonparametric regression estimators. *Journal of the American Statistical Association* 86(415), 665–72.

- Heckman, J. (1976). The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* 5, 475–92.
- Nelson, F. (1977). Censored Regression Models with Unobserved, Stochastic Censoring Thresholds. *Journal of Econometrics* 6, 309–28.
- Newey, W.K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R.F. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, pp. 2111–245. Amsterdam: North Holland.
- Reece, W. (1979). Charitable contributions: New evidence on household behavior. *American Economic Review* 69, 142–51.
- Romano, J.P. and A.F. Siegel (1986). *Counter Examples in Probability and Statistics*. Monterey, CA: Wadsworth.
- Schervish, M.J. (1995). *Theory of Statistics*. New York: Springer Verlag.
- Sheather, S.J. and M.C. Jones (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society Series B* 53, 683–90.
- Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica* 26, 24–36.
- Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. Boston, MA: The MIT Press.
- Zuehlke, T. (2003). Estimation of a Tobit model with unknown censoring threshold. *Applied Economics* 35, 1163–9.