INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION
PAPER

2014/10

A note on the Tobit model in
the presence of a duration variable

HAFNER, C. AND A. PREMINGER

A note on the Tobit model
in the presence of a duration variable

Christian M. Hafner[*][1] and Arie Preminger[2]

[1]*Université catholique de Louvain*
[2]*Ben Gurion University*

March 5, 2014

Abstract

The Tobit model (censored regression model) is an important basic model appearing in many applications in economics. In this paper we consider a duration Tobit model in which a duration variable which counts the number of times the data is being censored is included as a covariate. We show that in this case, the dependent variable eventually becomes degenerate, which makes the asymptotic Fisher information matrix singular, rendering the standard methods of asymptotic inference inapplicable. We provide a simulation study and an empirical application to support our results.

Keywords: Limited dependence, censoring, duration, labor supply,

JEL Classification: C24, J64

---

[*]Corresponding author, Institute of statistics, biostatistics and actuarial sciences, and CORE, Université catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium, email: `christian.hafner@uclouvain.be`

1 Introduction

Recently, Basu and de Jong (2009) have shown that including duration variables in binary choice regressions may be problematic and, in some situations, invalidate statistical inferences based on the asymptotic distribution of maximum likelihood estimators. For example, Frederiksen et al. (2007) use duration variables in several limited dependence models with group level heterogeneity in a panel data setting with large cross section and small time dimension. Extending this model to the case with large time dimension will lead to invalid inference based on asymptotic theory. The intuitive reason for the failure of asymptotic theory is that duration variables act as stochastic trends, such that the dependent variable may become degenerate.

In this paper, we show that a similar result holds for the classical regression model with censoring, often called the Tobit model. Apart from the model being Tobit rather than Probit, our paper is distinct from Basu and de Jong (2009) in several respects. First, Basu and de Jong (2009) assume that the errors are i.i.d. while we allow for heterogenous distributions and, hence, the inclusion of regressors other than the intercept. Second, they assume that the support of the error is unbounded. This requires adding a uniform integrability assumption on the error term which is not needed if the errors are bounded from above. Finally, we provide a simulation study and a real data example for our results.

We show under mild conditions on ε_t that the inclusion of a duration variable in the model causes the dependent variable to converge to a degenerate limit distribution, which makes the Fisher information matrix singular and asymptotic statistical inference impossible.

The following section presents the model, the assumptions and main results. Section 3 reports a simulation study to support the theoretical results, and Section 4 illustrates the effects of including a duration variable in Tobit-type models in a well-known textbook example. Section 5 concludes.

2 The model and main results

Let us consider the following censored regression model:

$$y_t = \max(0, \gamma_0 z_t + x_t' \beta_0 + \varepsilon_t) \tag{1}$$

where $\{\varepsilon_t\}$ has mean zero and finite variance σ_0^2, x_t is a $k \times 1$ vector of regressors and β_0 is a $k \times 1$ vector of parameters of interest. The regressors x_t are assumed to be sequences of

independent but not necessarily identically distributed random variables. In addition, the model includes a duration variable z_t which is defined as the number of consecutive zeros of the y_t sequence leaving out the current period. Thus z_t counts the number of times the data is being censored which may have some economic meaning. See e.g. Frederiksen et al. (2007) whose empirical application uses limited dependence models with group level heterogeneity in a panel framework. For example, y_t could measure the supply of working hours which is observed only if a person is employed, where z_t measures the duration of unemployment. In our setting we assume that $\gamma_0 < 0$ which is motivated by the assumption that the longer a person is absent from work activity, the smaller will be his/her working hours supply.

Let $F_\varepsilon^t(\cdot) > 0$ denote the distribution of ε_t and $I(\cdot)$ the indicator function. Our results are given under two separate sets of assumptions. The first set of assumptions is given in the following.

(A1) $\{\varepsilon_t\}$ is a sequence of independent random variables.

(A2) $\inf_t F_\varepsilon^t(z) > 0, \ \forall z \in \mathbb{R}$.

(A3) $\sup_t \varepsilon_t \leq M < \infty$.

In the proofs we use the simplified model $y_t = \max(0, \gamma_0 z_t + \varepsilon_t)$. However, since we do not impose a centering assumption on ε_t and allow for independent heterogeneous distributions for the errors, the results remain true for model (1), which may include other stochastic independent regressors. Assumption (A3) imposes an upper bound for the support of ε_t, which allows to obtain a strong convergence result for the y_t sequence.

Theorem 1 *Under assumptions (A1)-(A3),* $\lim_{t \to \infty} y_t = 0$ *almost surely.*

This theorem shows that, under the first set of assumptions, the sequence y_t converges almost surely to zero. Note that this result is stronger than the analogous convergence result of Basu and de Jong (2009) for the probit model, which is in probability.

The second set of assumptions, used for the following theorem, is without Assumption (A3), so that the support of ε_t is unbounded as in Basu and de Jong (2009). In this case, however, for the tobit model, we need the following additional assumption, which is a uniform integrability condition.

(A4) $\limsup_{M \to \infty} \sup_t E\left(|\varepsilon_t|^{1+\delta} I(\varepsilon_t \geq M) \right) = 0$ for some $\delta > 0$.

We now have the following result.

Theorem 2 *Under assumptions (A1), (A2) and (A4), $\lim_{t \to \infty} y_t = 0$ in probability.*

The intuition behind the convergence results of both theorems is as follows. Since there is a positive probability that the error will be negative, the y_t sequence will eventually be zero in some period. This implies that the probability that y_t will continue to be zero in the next period increases because z_t increases by one. As a result, the probability that $\varepsilon_t > |\gamma_0 z_t + x_t' \beta_0|$ will approach zero causing the y_t sequence to get stuck at zero.

We proceed by examining the implications of these results for the maximum likelihood estimator (MLE) of the model parameters. We assume for simplicity and without loss of generality that only the duration variable z_t is included in the model and that $\sigma_0 = 1$. In order to formulate the likelihood function we note that the Tobit model can be written as

$$y_t = D_t(\gamma_0 z_t + \varepsilon_t) \tag{2}$$

where $D_t = I(\gamma_0 z_t + \varepsilon_t > 0)$ and $z_t = \sum_{j=1}^{t} \prod_{i=1}^{j} (1 - D_{t-i})$. Next, by assuming that the error term is Gaussian we get

$$f(y_t | y_{t-1}, \ldots, y_0; \theta) = [\phi(y_t - \gamma z_t)]^{D_t} [\Phi(-\gamma z_t)]^{1-D_t}$$

Hence the conditional likelihood for the sample can be written as

$$\log L_T(\gamma) = \frac{1}{n} \sum_{t=1}^{n} D_t \log \phi(y_t - \gamma z_t) + (1 - D_t) \log \Phi(-\gamma z_t) \tag{3}$$

where (3) is maximized by $\hat{\gamma}_n$ and the Hessian conditional on z_t is given by

$$E(H_t(\gamma_0) | z_t) = -\{[1 - \Phi(-\gamma_0 z_t)] + \Phi(-\gamma_0 z_t) \lambda(z_t) [z_t + \lambda(z_t)]\} z_t^2, \tag{4}$$

where $\lambda(z_t) = \phi(\gamma z_t)/(1 - \Phi(\gamma z_t))$ which is known as the inverse Mills' ratio.

By standard asymptotic results for MLE, we would expect that

$$\sqrt{n}(\hat{\gamma}_n - \gamma) \overset{a}{\sim} N\left(0, -\left\{\tfrac{1}{n} \sum_{t=1}^{n} E[H_t(\gamma_0)]\right\}^{-1}\right).$$

However, Theorems 1 and 2 imply that $z_t \to \infty$, and by applying the l'Hôpital's rule and the dominated convergence theorem to (4), it follows that $E(H_t) \to 0$ as $t \to \infty$. Therefore, the Fisher information matrix is singular, and estimation and statistical inference using MLE is not possible for the Tobit model with duration variable.

4

n	$\gamma = -0.1$	$\gamma = -0.15$	$\gamma = -0.2$
	$\beta = 1$		
100	-0.35	-0.18	-0.048
500	-0.21	-0.006	-0.0001
2000	-0.03	0.00	0.00
10000	0.00	0.00	0.00
	$\beta = 2$		
100	-0.55	-0.43	-0.26
500	-0.50	-0.18	-0.02
2000	-0.39	-0.006	0.00
10000	-0.09	0.00	0.00

Table 1: *Monte Carlo mean of y_n over $K = 10,000$ replications for alternative coefficients of the duration variable γ, regression coefficients β, and sample sizes n.*

3 Simulation study

In the following we present results of a simulation study. For each set of parameters, we generate $K = 10000$ processes $\{y_t\}$ of length n, given by

$$
\begin{aligned}
y_t &= D_t(\gamma_0 z_t + \beta x_t + \varepsilon_t) \\
D_t &= I(\gamma_0 z_t + \beta x_t + \varepsilon_t > 0) \\
z_t &= \sum\nolimits_{j=1}^{t} \prod\nolimits_{i=1}^{j} (1 - D_{t-i}).
\end{aligned}
$$

The regressor x_t is assumed to be i.i.d. uniformly distributed on $(-1, 1)$, while the error $\varepsilon_t \sim N(0, \sigma^2)$, independent of x_t. For this specification, $\beta x_t + \varepsilon_t$ is i.i.d. with mean zero and variance $\sigma^2 + \beta^2/3$. The parameters β and σ essentially determine the dispersion of $\beta x_t + \varepsilon_t$, so we fix one of them, $\sigma = 1$, and let $\beta \in \{1, 2\}$ to see how the effect of the duration variable depends on the dispersion of the regression plus error term. For the sample size n, we choose $n \in \{100, 500, 2000, 10000\}$. Table 1 reports the mean of y_n over the K replications.

The general conclusion is that in every case, the Monte Carlo mean of y_n converges to zero as n increases. The same is true for the corresponding standard deviations (not reported here to economize on space). The effect is stronger when γ, the coefficient of the duration variable, is increasing in absolute value. Convergence to zero takes longer when

β and hence the dispersion of $\beta x_t + \varepsilon_t$ is larger, but the conclusions are the same.

4   Empirical example

We consider the classical textbook example of female labor supply based on Mroz (1987) and data of the 1975 panel survey of income dynamics, see e.g. Example 16.3 of Wooldridge (2002). The data set contains 753 observations on married white women, aged 30-60. Of the full sample, 428 women had positive working hours, while the remaining 325 women did not work in 1975. First, a tobit model is estimated as in Wooldridge (2002), where working hours is the dependent variable, which is left-censored at zero. We follow Wooldridge in specifying the explanatory variables: a constant, non-wife income, education, work experience, work experience squared, age, number of kids younger than 6, and number of kids 6 to 18 years.

   We then introduce a duration variable z_t, defined by the number of successive periods with zero working hours for a given woman. Since we do not have data for more than one time period, we have to make assumptions about the impact of duration on working supply. It is economically reasonable to assume that the impact is negative: the longer a woman is absent from work activity, the smaller will be her working hours supply. Thus, alternative negative coefficients γ are selected for the duration impact. The descriptive statistics of the dependent variable are the following: Mean and standard deviation are given by 740 and 871, respectively, while minimum and maximum are zero and 4950, respectively. Based on these statistics, we choose a range of -100 to -500 for the γ parameter. We choose the average profile for each one of the explanatory variables, and then simulate the model

$$y_t = \max(0, \gamma z_t + x_t'\beta + \varepsilon_t), \quad t = 1, \ldots, n$$

for alternative values of n and γ. The number of replications is again $K = 10,000$. Table 2 reports the mean and standard deviation of y_n over the K replications. Choosing a sufficiently large absolute value of γ leads to degenerated y_t series, which invalidates inference based on asymptotic theory.

   To assess how the maximum likelihood parameter estimates change when including a duration variable, we make the simplifying assumption that the explanatory variables remain unchanged over time for all women, except for the duration variable. We simulate the model, which was estimated without duration ($\gamma = 0$), $K = 10,000$ times with

n	$\gamma = -100$	$\gamma = -200$	$\gamma = -300$	$\gamma = -400$	$\gamma = -500$
100	558 (731)	253 (555)	34 (210)	2.45 (47)	0.05 (9.5)
200	555 (723)	126 (427)	1.5 (40.3)	0 (0)	0 (0)
500	518 (718)	11 (139)	0 (0)	0 (0)	0 (0)
1000	486 (695)	0.05 (11.5)	0 (0)	0 (0)	0 (0)

Table 2: *Empirical example: Means of simulated y_n over $K = 10,000$ replications for the estimated tobit model including a duration variable. Corresponding standard deviations are in parentheses.*

$n = 100$, including the duration variable and alternative values for γ. For each simulated data series, we re-estimate the parameter vector β by MLE. The following table reports the results. Not surprisingly, we observe dramatic changes in parameter estimates. This confirms our message when including duration variables in models for limited dependent variables.

5 Conclusions

We have shown that including duration variables in censored regression models may lead to degenerate limited dependent variables and hence invalidate asymptotic inference. If excluding such variables from the regression is not an option, then one might bound or transform them in such a way that the stochastic trend character is avoided.

Appendix

Throughout the Appendix let $\Delta_{in} = \int_n^{+\infty} |x| dF_\varepsilon^i(x)$ and "i.o." stand for infinitely often. Also we assume without loss of generality that $\gamma_0 = -1$.

For the original process $\{y_t : t \geq 0\}$ we define the series $\{\pi_n\}$ of stopping times, as follows:

$$\pi_n = \inf \left\{ t : \cap_{i=0}^{n-1} \{y_{t-i} = 0\} \right\}$$

where π_n is the first time that the series has generated a realization of n consecutive zeros. Note that π_n is a random variable defined on the same probability space as the original process, taking values in the time set $\mathbb{N} = \{1, 2, ...\}$.

coeff	$\gamma = 0$	$\gamma = -100$	$\gamma = -200$	$\gamma = -300$	$\gamma = -400$	$\gamma = -500$
C	965.3053	1020.666	968.3520	953.6852	870.3407	725.8959
WE	80.64561	69.80077	67.16829	58.47226	53.64403	52.25296
PRIN	-0.008814	-0.008979	-0.009716	-0.008809	-0.007575	-0.005903
WA	-54.40501	-54.29680	-53.64290	-53.10886	-51.15087	-50.51516
KL6	-894.0217	-910.2960	-955.0989	-963.9229	-913.5043	-882.4327
K618	-16.21800	-20.77361	-28.27405	-37.84909	-43.12263	-34.47221
AX	131.5643	132.8219	131.5351	132.3745	127.5778	126.9951
AX2	-1.864158	-1.929180	-1.910915	-1.938532	-1.842695	-1.805661
R^2	0.273	0.266	0.259	0.251	0.245	0.235

Table 3: *MLE parameter estimates of β for alternative values of γ. Variables are: a constant (C), non-wife income (PRIN), education (WE), work experience (AX), work experience squared (AX2), age (WA), number of kids younger than 6 (KL6), and number of kids 6 to 18 years (K618).*

Lemma 1 π_n *is a well defined random variable.*

Proof: Note first that

$$
\begin{aligned}
P\left(\pi_n \le t+n \mid \pi_n > t\right) & \ge & P\left(\cap_{i=1}^n \{y_{t+i} = 0\} \mid \pi_n > t\right) \\
& = & P\left(\cap_{i=1}^n \{y_{t+i} = 0\}\right) \\
& = & \prod_{i=1}^n F_\varepsilon^{t+i}(i-1) \ge \eta > 0.
\end{aligned}
\tag{5}
$$

The first inequality follows because the event $\cap_{i=1}^n \{y_{t+i} = 0\}$ is contained in the set of events $\pi_n \le t+n$, and the first equality follows because $\cap_{i=1}^n \{y_{t+i} = 0\}$ is independent of y_j, $j = 1, \dots, t$.

For some η, the last inequality in (5) follows from assumption A2. Thus, for all $t \in \mathbb{N}$,

$$
\begin{aligned}
P(\pi_n > t+n) & = & P(\pi_n > t) - P(t < \pi_n \le t+n) \\
& = & P(\pi_n > t) - P(\pi_n > t, \ \pi_n \le t+n) \\
& \le & P(\pi_n > t)(1-\eta)
\end{aligned}
\tag{6}
$$

Iterating on multiples of n yields, for every $k \ge 0$,

$$
P(\pi_n > kn) \le (1-\eta)^k \to 0 \quad \text{as } k \to \infty,
\tag{7}
$$

8

and since $P(\pi_n = \infty) \leq P(\pi_n > kn)$ we deduce that $P(\pi_n = \infty) = 0$. Hence π_n is almost surely finite. $\square$

Lemma 2 *Under assumption (A4), $\Delta_{in} = o(n^{-\delta})$ for some $\delta > 0$.*

Proof: Note that

$$E\left(|\varepsilon_i|^{1+\delta} I(\varepsilon_i \geq n)\right) \geq n^\delta E\left(|\varepsilon_i| I(\varepsilon_i \geq n)\right) = n^\delta \Delta_{in} \tag{8}$$

The result follows by letting $n \to \infty$, since the LHS approaches zero by assumption (A4). $\square$

Proof of Theorem 1: The proof uses similar arguments as in Theorem 1 of Basu and de Jong (2009). Define $I_n = \{j \in \mathbb{N} : P(\pi_n = j) > 0\}$. Let $\bar{P} = \sup_{j \in I_n} P(y_t > 0 \ i.o. | \pi_n = j)$ and

$$\bar{j} = \inf\left\{j : j \in I_n, \ P(y_t > 0 \ i.o| \pi_n = j) = \bar{P}\right\} \tag{9}$$

Note that

$$P\left(\cup_{j \in I_n}\{\pi_n = j\}\right) = \sum_{j \in I_n} P(\pi_n = j) = 1 \tag{10}$$

Then,

$$
\begin{aligned}
P(y_t > 0 \ i.o) \ &= \ P\left(\{y_t > 0 \ i.o.\} \cap \{\cup_{j \in I_n}\{\pi_n = j\}\}\right) \\
&= \ P\left(\cup_{j \in I_n}\{\{\pi_n = j\} \cap \{y_t > 0 \ i.o.\}\}\right) \\
&\leq \ \sum_{j \in I_n} P\left(\{\pi_n = j\} \cap \{y_t > 0 \ i.o.\}\right) \\
&= \ \sum_{j \in I_n} P(y_t > 0 \ i.o.| \pi_n = j) P(\pi_n = j) \\
&\leq \ P(y_t > 0 \ i.o.| \pi_n = \bar{j}) \sum_{j \in I_n} P(\pi_n = j) \\
&= \ P(y_t > 0 \ i.o.| \pi_n = \bar{j}) \\
&\leq \ P(\exists t \geq \bar{j}, \ y_t > 0| \pi_n = \bar{j}) \\
&\leq \ P\left(\cup_{l \geq n}\{\varepsilon_{\bar{j}+l-n} - l \geq 0\}\big| \pi_n = \bar{j}\right) \\
&= \ \sum_{l=n}^{\infty} P\left(\varepsilon_{\bar{j}+l-n} \geq l\right) \tag{11}
\end{aligned}
$$

The first equality follows from (10), the first inequality holds by Boole's inequality. The conditional probability in the third equality is well defined since $j \in I_n$. The second inequality follows from (9). The last equality follows from Assumption (A1) and because $\{\pi_n = \bar{j}\}$ depends only on $\{y_0, \varepsilon_1, \ldots, \varepsilon_{\bar{j}}\}$. Now, by Assumptions (A3) there exists $M > 0$ such that $\varepsilon_{\bar{j}+l-n} < M$, hence the last term equals zero as $n \to \infty$. $\square$

Proof of Theorem 2: Without loss of generality we can assume that $\bar{j} = 0$. When the support of ε_t is unbounded we have

$$
\begin{aligned}
\sum_{l=n}^{\infty} P\left(l \leq \varepsilon_{l-n}\right) &= \sum_{l=n}^{\infty} \sum_{i=l}^{\infty} P\left(i \leq \varepsilon_{l-n} \leq (i+1)\right) \\
&= \sum_{i=n}^{\infty} \sum_{j=0}^{i-n} \int_{i}^{(i+1)} dF_{\varepsilon}^{j}(x) \\
&\leq \sum_{i=n}^{\infty} \sum_{j=0}^{i-1} \int_{i}^{(i+1)} \frac{|x|}{i} dF_{\varepsilon}^{j}(x) \\
&\leq \frac{1}{n} \sum_{j=0}^{n-1} \int_{n}^{+\infty} |x| dF_{\varepsilon}^{j}(x) + \sum_{j=1}^{\infty} \frac{1}{n+j} \int_{(n+j)}^{+\infty} |x| dF_{\varepsilon}^{n+j}(x) \\
&= \frac{1}{n} \sum_{j=0}^{n-1} \Delta_{jn} + \sum_{j=1}^{\infty} \frac{\Delta_{(n+j)}}{n+j} \\
&\leq o(n^{-\delta}) + \int_{n}^{+\infty} \frac{1}{(n+j)^{1+\delta}}
\end{aligned}
\tag{12}
$$

where the last inequality follows from Lemma 2 and the last expression converges to zero as $n \to \infty$. The desired result follows from (11). □

References

Basu, D. and de Jong, R. (2009). A Note on Binary Choice Duration Model, *Economics Letters*, 102(1),17–18.

Frederiksen, A., Honoré, B. and Hu, L. (2007). Discrete Time Duration Models with Group-level Heterogeneity. *Journal of Econometrics* 141 (2007) 1014–1043

Mroz, Th. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions, *Econometrica*, 55, 765-799.

Wooldridge, J. (2002). Econometric analysis of cross section and panel data, MIT Press.