AN EXTENSION OF THE NORMAL CENSORED REGRESSION MODEL. ESTIMATION AND APPLICATIONS

By

Karla Andrea Muñoz Gajardo

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
AT
PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
SANTIAGO, CHILE
AUGUST 14 2009

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
DEPARTMENTO DE
ESTADÍSTICA

The undersigned hereby certify that they have read and recommend to the Facultad de Matemáticas for acceptance a thesis entitled "**An Extension of the Normal Censored Regression Model. Estimation and Applications**" by **Karla Andrea Muñoz Gajardo** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**.

Dated: August 14 2009

External Examiner: ______________________
Heleno Bolfarine

Research Supervisor: ______________________
Reinaldo Arellano-Valle

Examing Committee: ______________________
Wilfredo Palma

____
Hugo Salinas

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

Date: **August 14 2009**

Author: **Karla Andrea Muñoz Gajardo**

Title: **An Extension of the Normal Censored Regression Model. Estimation and Applications**

Department: **Estadística**

Degree: **Ph.D.** Convocation: **August** Year: **2009**

Permission is herewith granted to Pontificia Universidad Católica de Chile to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions.

____

Signature of Author

The astonishing thing is when you are most truly alone, when you truly enter a state of solitude , that is the moment when you are not alone anymore, when you start to feel your connection with others. In the process of writing or thinking about yourself, you actually become someone else.

PAUL AUSTER
The Art of Hunger

Table of Contents

Acknowledgements

I would like to thank Reinaldo Arellano-Valle, my supervisor, for his many suggestions, constant and unconditional support, as a teacher during this research and also as a good friend.

Thank you so much Professor Fernando A. Quintana. He taught me almost everything i know, which is discipline, perseverance, and the particular motivation expressed in his classes.

Thanks to my friends, for their permanent support and sincere loving. I have to mention them: Tomás Jesús Beltrán, Francisco Kúncar, Andrés Valdivia, María Paz Uribe, Ximena Saavedra, Juan Carlos Mendoza, Paola Armijo, and Luz María Olivera. I have the pleasure of meeting several people that are always in my heart: Jorge González, Fanny Ordóñez, Mauricio Castro, Daniela Rowlands and Jimena Mendoza; wonderful people, who had nothing to do with this research, but make me constantly a better person. They are the closest to what's called best friends.

I should also mention that my graduate studies in Chile were supported in part by the DIPUC, VRAID (Vicerrectoría Adjunta de Investigación y Doctorado). FONDECYT (N° 1085241), for making possible the assistance to some important meetings (invited sessions, oral communications).

Of course, I am grateful to my parents and family for their patience and love. Without them this work would never have come into existence (literally).

Finally, I wish to thank the following: Macarena (for her wisdom and faithful friendship); Raúl, Marcela, Paula and Claudia (for changing my life from worse to better); Gaby and Natalia (for the painful introduction to the Dirichlet processes team); Paula, Mallén, Vicky, Carolina,(for all the good and bad times we had together); under and graduate students (for teaching me i have to go easy and slow); Jacques, Alain, Patricia and Lía (they are definitely the sweetest thing); Oasis (they know why); and my little niece Lou (we simply love each other).

Santiago Karla A. Muñoz Gajardo
August 2009

Abstract

Tobit models are regression models where the dependent variable is censored (limited). These models are very common , e.g., in econometric analysis, and most of the literature related to this kind of models is developed under the normal assumption.

A brief review on Tobin's work and some theoretical results for the limited dependent normal regression model are presented. This article takes up a linear model as the one described above, in which the disturbances are independent and have identical Student-t distribution. In the context of maximum likelihood estimation, we provide an expression to the information matrix, under a convenient re-parametrization of the model.

In this thesis we discuss the possibility of extending the Tobit model through some new distributional assumptions. Some parametrical aspects about the model are also presented in the sense of showing how the estimation can be made. The inference and all the asymptotical results are compared with those described in the literature, for instance in Amemiya (1984), Olsen (1978), among some others.

An interesting way of extending Tobin's model is supposing that the statistical distribution of the perturbations are not normal any longer, so we can assume some other distribution like the Student-t. The importance of choosing such distribution is based on the robustness that the Student-t posses; the t distribution provides a useful extension of the normal for statistical modeling of data sets involving errors with longer than normal tails. The degrees of freedom parameter of the t distribution provides a convenient dimension for achieving robust statistical inference, with moderate increases in computational complexity for many models. Basically, the thesis

describes the censored model using a different approach to the normal model, which is pretty much known that is quite vulnerable to the presence of outliers. Furthermore, many models treat this problem editing the data that represents outlying observations; this last procedure may cause the fact that the uncertainty could not be well reflected when the inference is made.

In addition, a Monte Carlo simulation study is made in order to adjust the non normal censored regression model, meaning, under the assumption of Student-t perturbations, and in addition compare the estimations as the degrees of freedom increase. It's intuitive thinking that there will be similarities in the normal case and the Student-t case, but the difference is that the last model is more capable to reach those extreme values in the data, making it more robust than the normal model. This feature has the purpose to convince the reader that it is feasible to extend Tobit models through a distributional constrain; moreover, estimation of the information matrix, although it takes lot of analytical computations, is available considering the parameter ν (which represents the degrees of freedom that the model poses) unknown. These calculations are extremely large and exhaustive, but worthy for the main purpose.

Chapter 1

Introduction

> *"What do you mean, less than nothing?" replied Wilbur. "I don't think there is any such thing as less than nothing. Nothing is absolutely the limit of nothingness. It's the lowest you can go. It's the end of the line. How can something be less than nothing? If there were something that was less than nothing then nothing would not be nothing, it would be something-even though it's just a very little bit of something. But if nothing is nothing, then nothing has nothing that is less than it is."*

E. B. White, *Charlotte's Web*
(New York: Harper, 1952) p. 28.

1.1 Limited Dependent Variables in Regression Analysis.

A crucial aspect of any empirical research is to develop a research design to understand the phenomenon of interest and to guide the selection of an appropriate statistical method. A first step toward the choice of statistical method is deciding what measure of the dependent variable can best represent the concept of interest. To arrive at the appropriate measure, the researcher will need to determine the range of variation of the phenomenon of interest, the nature of its distribution, and how fine or gross to make the distinction between particular attributes of the phenomenon. Jointly with the purpose of the research, these are the considerations that drive the final choice of measure for the dependent variable. It is essential that the dependent variable be well-measured, well-distributed, and have enough variance so that there is indeed something to explain. At the same time, we define a model as a simplified description of reality that is at least potentially useful in decision making. Since models are simplified, they are never literally true: whatever the ”data generating process” may be, it is not the model. Since they are constructed for the purpose of decision making, different decision problems can appropriately lead to different models despite the fact the reality they simplify is the same. All scientific models have certain features in common. One is that they often reduce an aspect of reality to a few quantitative concepts that are unobservable but organize observable in a way that is useful in decision making. In this context, we'll say that a censored dependent variable occurs when the values of the measured phenomenon, above or below some threshold value, are all assigned the same value.

The problem of estimation for a regression model where the dependent (or endogenous) variable is limited, has been studied in different fields: econometric analysis, clinical essays, wide range of political phenomena, among some others. In econometric, some of these models are commonly classified as: truncated regression models, discrete regression models, and censored regression models. The main goal of this thesis is to focus on the study of the censored regression model, making a wide review based on the common assumptions, and also introducing a different perspective that will allows us to obtain new interest quantities in the statistical inference issue.

As a first motivating example, let's analyze the problem of mobile phones consumption in a particular city. If an investigator is interest in this phenomenon, he needs to incorporate information about people who has a mobile phone. A person's needs in the last years often includes a communication media, even though a person's incoming are not enough. Should the investigator put such person out of his research? The answer is commonly no, because the "non-mobile-person" is also an integral part of the information, so instead of truncating the data we are censoring.

Most of the results on the normal censored regression model are based on the development for the probit model, where the variable of theoretical interest, let's say, y^*, is unobserved. Instead, we may observe a dummy variable, represented by 1 if it is observed and 0 in other case. So, since James Tobin's work (1958), the hybrid of probit and regression analysis makes this kind of model be nicknamed Tobit model. Specifically, we consider the truncation where the threshold is the zero, hence our Tobit model accept those strictly positive observations. On his example, Tobin refers to the relationship between the households incomes and the households expenditure, on various categories of goods. The zero expenditure on some luxury goods reports a low level of income. Compare with those households who in fact made that kind of expenditure, the range of variability of the observe expenditure is wide, meaning that there's a group of observations concentrated around zero; besides, this reasoning implies that there cannot be negative expenditures. That is, for some observations the observed response is not the actual response, but rather the censoring value (zero), and an indicator that censoring (from below) has occurred.

As we know, the limited response is directly related with the disturbances distribution, which typically is assume to be normal with common variance. Our primary focus is to extend this assumption to the case where the error terms are Student-t distributed, describing the likelihood function to find the estimators for the regression parameters and of course to obtain the information matrix, that conduces us to the standard errors of estimations.

A *limited dependent variable*, y, is, as its name suggests, one whose set of possible values is restricted in some obvious way. *Limited dependent variable models* have traditionally been estimated using the method of maximum likelihood, which naturally requires the specification of a parametric likelihood function. The idea of ”censoring” is that some data above or below the threshold are misreported at the threshold, so the observed data are generated by a mixed distribution with both a continuous and a discrete component. The censoring process may be explicit in the data collection process, or it may be a by-product of economic constraints involved in constructing the data set.

The first censored regression model developed by Tobin in 1958 explains the consumption of durable goods. Tobin observed that for many households, the consumption level, i.e. the purchases, in a particular period was reported as zero. Such model, specifies that the latent, or ideal, value of consumption may be negative, which means that the households would prefer selling over buying. All that can be reported is that the household purchased zero units of the good in question. Further examples of limited dependent variables are: zero expenditure and corner solutions: labor force participation, smoking, demand, among some others.

A regression model is said to be *censored* when the recorded data on the dependent variable (the response) cuts off outside a certain range with multiple observations at the endpoints of that range. When the data are censored, variation in the observed dependent variable will understate the effect of the regressors on the "true" dependent variable. As a result, standard ordinary least squares regression using censored data will typically result in coefficient estimates that are biased toward zero. Traditional statistical analysis uses maximum likelihood or related procedures to deal with the problem of censored data. However, the validity of such methods requires the correct specification of the error terms distribution, which can be quite problematic in practice.

If the distribution of the error terms (or disturbances), given the regressors has a simple parametric form, for instance, normally distributed and homoscedastic errors, we can derive and maximize the likelihood function. Thus, it is important to develop estimation methods that provides consistent estimates for censored data even when the error distribution is non normal or heteroscedastic.

On the other hand, a dependent variable in a model is *truncated* if observations cannot be seen when it takes on values in some range. That is, both the independent and the dependent variables are not observed when the dependent variable is in that range. A natural example is when we have data on consumption purchases, if a consumer's willingness-to-pay for a certain product is negative, we will never see evidence of it no matter how low the price goes. Price observations are truncated at zero, along with identifying characteristics of the consumer in this kind of data. A model is called *truncated* if the observations outside of a specified range are totally lost. This kind of model will not be specifically study on this work, nevertheless some relevant properties about truncated distributions will be recalled.

1.1.1 Some examples of the censored regression model in economics.

Censored and truncated models have being used in different fields. For the *Tobit model* there has been various generalizations, which have been specially popular among economists, psychologists, and mostly social scientists. Goldberger (1964) named it *Tobit* because of its similarities to *probit* models.

Economists and sociologists have used this kind of models to analyze the duration of such phenomena as unemployment, welfare receipt, employment in a particular job, residing in a particular region, marriage, and the time between births.

Between 1958 and 1970, when Tobin's article appeared, the Tobit model was infrequently used in econometric applications, but since the early 1970's numerous applications ranging over a wide area of economics have appeared and continue to appear. This phenomenon is due to the recent increase in the availability of micro sample survey data which the Tobit model analyzes well and to the recent advances of technology that make possible the estimation of the model under a large scale data.

As a result of many methods of estimation, like maximum likelihood, moments, two steps estimators, use of latent variables, etc., the results indicate that iterative procedures are quite optimal in the sense that under some convenient reparametrization, such estimations are consistent and have desirable asymptotic properties. Some interesting papers have discussed this estimation problem. For instance, in the survey by Heckman (1979) and Singer (1976) the problem of estimation bias is mentioned on straight relationship with the estimators of the regression coefficients β and the scale parameter σ.

An early empirical example is described by Tobin (1958), who obtained the maximum likelihood estimates of his model applied to data on 735 non-farm households obtained from Surveys of Consumer Finances. The dependent variable of his estimated model was actually the ratio of total durable goods expenditure to disposable income and the independent variables were the age of the head of the household and the ratio of liquid assets to disposable income.

Further examples about censored regression models are the following:

(1) *Demand for durable goods.*

If we have a survey data on consumer expenditures, we find that most households report zero expenditures on automobiles or major household goods during a year. However, among those households that make any such expenditures, there will be wide variation in the amounts. Thus, there will be a lot of observations concentrated around zero. As mentioned above, Tobin analyzed this problem formulating the previous regression model taking as dependent variable the expenditures, y, and considering a set of explanatory variables, x.

(2) *Changes in holding of liquid assets.*

Consider a change in a household's holding of liquid assets during a year. This variable can be positive or negative, but it can´t be smaller than the negative of the household's assets at the beginning of the year, basically because there is no chance of liquidate more assets that one owns. Note that here the threshold is not the zero and is different for different individuals

(3) *Married women in the labor force.*

A problem that has been analyzed by Nelson (1975, 1977) is the reservation wage of the housewife , named y^*, based on her valuation of time in the household. Let y be the market wage based on an employer's valuation of her effort. The woman participates in the labor force if $y > y^*$. Otherwise, she does not. In any given sample we have observations on y for those women who participate in the labor force, and we have no observations on y for those who do not. For these women we know only that we know only that $y^* \geq y$. Given these data, we have to estimate the coefficients in the equations explaining y^* and y, as in the following model:

$$y_i = \begin{cases} x_i^T\beta + u_i \\ z_i^T\gamma + v_i. \end{cases}$$

1.1.2 Normal censored and truncated variables.

In this section we shall consider location-scale normal models in which the variable of interest is limited by truncation or censoring. We shall first explain the difference between truncated and censored samples and then consider the analysis of this problem in the linear models context.

Suppose y^* has a normal distribution, with mean μ and variance σ^2. Suppose we consider a sample of size N $(y_1^*, y_2^* \ldots, y_N^*)$ and record only those values of y^* greater than a constant c. For those values of $y^* \leq c$ we record only the value c. The observations are

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > c \\ c & \text{otherwise} \end{cases},$$

with $i = 1, \ldots, N$.

The resulting sample $y_1, y_2, \ldots, y_N$ is said to be a *censored sample*. Note that, the y_i has the same distribution of the y_i^* when $y_i^* > c$. For the observations $y_i = c$, all we know is $P(y_i = c) = P(y_i^* \leq c)$.

Hence, the likelihood function for estimation of the parameters μ and σ^2 is

$$L(\mu, \sigma^2 \mid y_1, \ldots, y_N) = \prod_{y_i^* > c} \frac{1}{\sigma} \phi \left(\frac{y_i - \mu}{\sigma} \right) \prod_{y_i^* \leq c} \Phi \left(\frac{c - \mu}{\sigma} \right)$$

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

Suppose now that before the sample is drawn we truncate the distribution of y^* at the point $y^* = c$, so that no observations are drawn for $y^* > c$. Under these conditions, the pdf of the truncated normal distribution from which the samples are drawn is

$$f(y^* \mid y^* < c) = \frac{1}{\sigma}\phi\left(\frac{y^* - \mu}{\sigma}\right) \Big/ \Phi\left(\frac{c - \mu}{\sigma}\right), \quad -\infty < y^* \leq c,$$

where $\Phi[(c - \mu)/\sigma]$ is the normalizing constant, because it is the integral of the numerator over the range $-\infty < y^* \leq c$.

A sample from this truncated normal distribution is called a *truncated sample.* For example, a sample drawn from families with incomes less than, say, $200,000.

In practice we can have samples that are doubly truncated, doubly censored, truncated-censored, and so forth. As an example of a truncated-censored sample, consider truncation at the level c_1 and censoring at level c_2 ($c_2 < c_1$); that is, only samples of y^* with $y^* \leq c_1$ are drawn, and among these samples only values of $y^* > c_2$ are recorded. For those observations $y^* \leq c_2$, we record c_2, that is, we observe

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > c_2 \\ c_2 & \text{otherwise.} \end{cases}$$

The likelihood function for this model is

$$L(\mu, \sigma^2 \mid y_1, y_2, \dots, y_N) \;=\; \left[\Phi\left(\frac{c_1-\mu}{\sigma}\right)\right]^{-N} \times \prod_{y_i^*>c_2} \frac{1}{\sigma}\phi\left(\frac{y_i-\mu}{\sigma}\right) \prod_{y_i^*\leq c_2} \Phi\left(\frac{c_2-\mu}{\sigma}\right)$$

One can consider further combinations of double truncation and double censoring, but the details are straightforward.

1.2 Tobit type model.

Since the use of econometric models with truncated or censored response variables came into increasing, it is important that the information they provide is fully and correctly used. One of these models that has an increasing use is the Tobit analysis, a model devised by Tobin (1958), in which it is assumed that the dependent variable has a number of its values clustered at a limited value, usually zero. For example, data for demand on consumption goods often have values clustered at zero; data on hours of work often have the same clustering. In other words, Tobit's type of models are censored regression models where the threshold that is considered is equal to zero. The Tobit technique uses all the observations, both those at the threshold and those above it, to estimate a regression line, and it is to be preferred, in general, over alternative techniques that estimate a line only with the observations above the threshold. In what follows, we understand by the Tobit model the censored regression model under the normality assumption.

The regression coefficients of the Tobit model frequently called the "beta" coefficients, provide more information than is commonly realized. In particular, it is well known that the Tobit type of modeling can be use to determine both changes in the probability of being above the threshold and changes in the value of the dependent variable if it is already above the threshold. This last decomposition can be quantified in rather useful and insightful ways.

The decomposition also has an important substantive economic and policy implications. For instance, in the question: how will the labor supply reduction induced by a negative income tax be spread between marginal decreases in hours worked and decreases the probability of working any hours?, it is not straightforward to know how to use the beta coefficients.

Now we write the Tobit model (or censored regression model) through $y_i = \max\{y_i^*, 0\}$, where $y_i^* = x_i^T\beta + u_i$, $i = 1, \ldots, N$; that is:

$$y_i = \begin{cases} x_i^T\beta + u_i, & \text{if } x_i^T\beta + u_i > 0 \\ 0, & \text{in other case.} \end{cases}$$

Here, the error terms u_i are assume to have $E(u_i \mid x_i) = 0$ and common scale parameter σ^2, hence

$$E(y_i \mid y_i > 0) = x_i^T\beta + \frac{\sigma f(z_i)}{F(z_i)}, \qquad z_i = x_i^T\beta/\sigma, \tag{1.1}$$

where f and F are the pdf and cdf of the distribution of u_i, respectively.

As shown in the review of McDonald and Moffitt (1979), the basic relationship between the expected value of any observation, $E(y_i)$, the expected value of y_i given $y_i > 0$, and the probability of been above the threshold, $F(z_i)$, is:

$$E(y_i) = F(z_i)E(y_i \mid y_i > 0). \tag{1.2}$$

The previous decomposition is achieved considering the effect produced by a change in the ith variable from x_i in y_i:

$$\frac{\partial E(y_i)}{\partial x_i} = F(z)\left(\frac{\partial E(y_i^*)}{\partial x_i}\right) + E(y_i^*)\left(\frac{\partial F(z)}{\partial x_i}\right). \tag{1.3}$$

Therefor, the total change in y_i can be disaggregated into two very intuitive parts: (1) the change of those y_i above the limit, weighted by the probability of being above the limit; and (2) the change in the probability of being above the limit, weighted by the expected value of y_i if above. The relatives magnitudes of these two quantities is an important indicator with substantive economic implications.

Assuming that β and σ have been already calculated, each of the term in equation (1.1) can be evaluated at some value of $x_i^T\beta$, usually on the mean of the $x_i's$, $\bar{x}$. The value of $E(y_i^*)$ can be calculated from equation (1.2), and the value of $F(z_i)$ can be obtained directly from statistical tables. The two partial derivatives are also presented:

$$\frac{\partial F(z_i)}{\partial x_i} = \frac{f(z_i)\beta_i}{\sigma} \tag{1.4}$$

and from equation (1.2):

$$\begin{aligned}\frac{\partial E(y_i^*)}{\partial x_i} &= \beta_i + \frac{\sigma}{F(z_i)}\frac{\partial f(z_i)}{\partial x_i} - \frac{\sigma f(z_i)}{F(z_i)^2}\frac{\partial F(z_i)}{\partial x_i} \\ &= \beta_i\left(1 - \frac{z_i f(z_i)}{F(z_i)} - \frac{f(z_i)^2}{F(z_i)^2}\right),\end{aligned} \tag{1.5}$$

using $F'(z_i) = f(z_i)$ and $f'(z_i) = -z_i f(z_i)$, for a standard normal density.

It should be noted from equation (1.3) that the effect of a change in x_i on $E(y_i^*)$ is not equal to β_i. It is a common error in the literature to assume that the Tobit beta coefficients measures the correct regression coefficients for observations above the threshold. As can be seen from equation (1.3), this is true only when x_i equals infinity, in which case $F(z_i) = 1$ and $f(z_i) = 0$. This will of course not hold at the mean of the sample or for any individual observation.

On the other hand, we should also note that when equations (1.2) and (1.3) are substituted into equation (1.1), the total effect $\partial E(y_i)/\partial x_i$ can be seen to equal simply $F(z_i)\beta_i$. Furthermore, by dividing both sides of equation (1.1) by $F(z_i)\beta_i$, it easily can be seen that the fraction of the total effect above the threshold, $\partial E(y_i^*)/\partial x_i$, is just $(1-z_i f(z_i)/F(z_i)-f(z_i)^2/F(z_i)^2)$. Thus, the information from the decomposition can be obtained calculating this fraction. In addition, this is also the fraction by which the β_i coefficients must be adjusted to obtain correct regression effects for observations above the limit.

Chapter 2

The Normal Tobit Model

"I remained for a while looking into the dark, this dense substance of darkness that had no bottom, which I couldn't understand. My thoughts could not grasp such a thing. It seemed to be a dark beyond all measurement, and I felt its presence weight me down. I closed my eyes and took to singing half aloud and rocking myself back and forth on the cot to amuse myself, but it did no good. The dark had captured my brain and gave me not an instant of peace. What if I myself became dissolved into the dark, turned into it?"

Paul Auster,*The Art of Hunger*

2.1 Introduction

In this chapter we give a complete description of the Tobit model, or so called the normal censored regression model, based on several works developed in the context of econometric analysis and other fields such as biostatistics and social sciences. Tobit models are refer to censored or truncated regression models in which the range of the dependent variable is constrained in some way. In economics, such a model was first suggested in a pioneering paper by Tobin (1958). He analyzed household expenditure on durable goods using a regression model that specifically took account of the fact that the expenditure (the dependent variable of his regression model) cannot be negative (see figure 2.1). Tobin called his model the model of *limited* dependent variables. It and its various generalizations are known popularly among economists as *Tobit models*, a phrase coined by Goldberger (1964), because of similarities to *probit models*. These models are also known as *censored* or *truncated* regression models. The model is called *truncated* if the observations outside a specified range are totally lost and *censored* if we can at least observe the exogenous variables. A more precise definition was given early in Chapter 1.

Censored and truncated regression models have been developed in other disciplines (notably, biometrics and engineering) more or less independently of their development in econometrics: Biometrician use the model to analyze the survival time of a patient. Censoring or truncation occurs when either a patient is still alive at the last observations date or he/she cannot be located. Similarly, engineers use the model to analyze the time to failure of material or of a machine or of a system. These model are also called *survival* or *duration models.* Sociologists and economists have also used survival models to analyze the duration of such phenomena as unemployment, welfare receipt, employment in a particular job, residence in a particular region, marriage, and the period of time between births. Mathematically, survival models belong to the same general class of models as Tobit models; survival models and Tobit models share certain characteristics.

Between 1958, when Tobin´s article appear, and 1970, the Tobit model was used infrequently in econometric applications, but since the early 1970s numerous applications ranging over a wide area of economics have appeared and continue to appear. This phenomenon increase in the availability of micro sample survey data, which the Tobit model analyzes well, and to a recent advance in computer technology that has made estimation of large scale Tobit models feasible. At the same time, many generalizations of the Tobit model and various estimation methods for these models have been proposed. In fact, models and estimation methods are now so numerous and diverse that it is difficult for econometricians to keep track of all the existing models and estimation methods and maintain a clear notion of their relative merits. Thus it is now particularly useful to examine the current situation and prepare a unified summary and critical assessment of existing results.

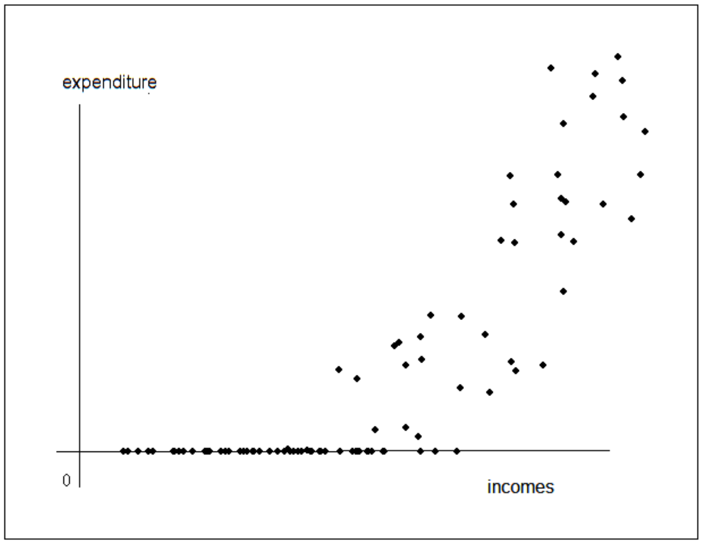

Figure 2.1: *An example of censored data*

2.2 The model specification.

The Tobit model is defined as follows

$$y_i = \begin{cases} x_i^T\beta + u_i & \text{if } x_i^T\beta + u_i > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

where β is a $k \times 1$ vector of unknown parameters; x_i is a $k \times 1$ vector of non stochastic constants; u_i are residuals that are independently and normally distributed, with mean zero and a common variance σ^2.

Our problem is to estimate β and σ^2 on the basis of N observations on y_i and x_i. This model was first studied by James Tobin (1958), and it's basically a censored normal regression model. As such, its estimation is related to the estimation of the censored and truncated normal distributions.

The literature on estimations of the parameters of the truncated normal distribution is extensive. However, it was Tobin who first discussed this problem in the regression context. Because he related his study to the literature on probit analysis, his model was nicknamed the tobit model (Tobin's probit) by Goldberger (1964). In this work, will be discussed the likelihood approach for the parameters estimations of the Tobit model.

Note that the threshold value in equation (2.1) is zero. This is not a very restrictive assumption, because if the model is

$$y_i = \begin{cases} x_i^T\beta + u_i & \text{if } y_i > c_i \\ c_i & \text{otherwise} \end{cases}$$

we can define

$$y_i^* = y_i - c_i, \quad x_i^* = \begin{pmatrix} x_i \\ c_i \end{pmatrix}, \quad \beta^* = \begin{pmatrix} \beta \\ -1 \end{pmatrix},$$

as the model in (2.1) applies with y, x, and β replaced by y^*, x^*, and β^*.

Similarly, for models with an upper constraint, so that

$$y_i = \begin{cases} x_i^T\beta + u_i & \text{if } x_i^T\beta + u_i < 0 \\ 0 & \text{otherwise} \end{cases}$$

we multiply y_i, x_i, and u_i by -1, and this reduces to the model in equation (2.1). Thus the Tobit model can be specified as in equation 2.1 without any loss of generality. Problems arise where the thresholds c_i are known only as stochastic functions of some other variables.

2.3 Maximum likelihood estimation.

For the model considered in equation (2.1), let N_0 be the number of observations for which $y_i = 0$, and N_1 the number of observations for which $y_i > 0$. Also, without any loss of generality, assume that the N_1 non zero observations for y_i occur first. For convenience, we define the following (notation here follows closely that of Amemiya (1973) and Fair (1977))

$$\begin{aligned}
\Phi_i &= \Phi\left(\frac{x_i^T\beta}{\sigma}\right) && (2.2)\\
&= \int_{-\infty}^{\frac{x_i^T\beta}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \quad t \in \mathbb{R},\\
\phi_i &= \frac{1}{\sigma}\phi\left(\frac{x_i^T\beta}{\sigma}\right) && (2.3)\\
&= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i^T\beta)^2/2\sigma^2},
\end{aligned}$$

where as was defined early, Φ and ϕ are, respectively, the cdf and pdf of the standardized normal distribution.

$$\begin{aligned}
\gamma_i &= \frac{\phi_i}{1-\Phi_i} && (2.4)\\
Y_1^T &= (y_1, y_2, \ldots, y_{N_1}) \quad \text{is a } 1 \times N_1 \text{ vector of } N_1 \text{ non zero observations on } y_i\\
X_1^T &= (x_1, x_2, \ldots, x_{N_1}) \quad \text{is a } k \times N_1 \text{ matrix of values of } x_i \text{ for non zero } y_i\\
X_0^T &= (x_{N_1+1}, \ldots, x_N) \quad \text{is a } k \times N_0 \text{ matrix of values of } x_i \text{ for } y_i = 0\\
\gamma_0^T &= (\gamma_{N_1+1}, \ldots, \gamma_N) \quad \text{is a } 1 \times N_0 \text{ vector of values of } \gamma_i \text{ for } y_i = 0. && (2.5)
\end{aligned}$$

For the observations y_i that are zero, all that we know is

$$P(y_i = 0) = P(u_i < -x_i^T\beta) = 1 - \Phi_i.$$

For the observations y_i that are greater than zero, we have

$$f(y_i \mid y_i > 0) \cdot P(y_i > 0) \;=\; \frac{1}{\sigma}\phi\left(\frac{y_i - x_i^T\beta}{\sigma}\right).$$

Hence the likelihood function is

$$L \;=\; \prod_0 (1 - \Phi_i) \prod_1 \frac{1}{\sigma}\phi\left(\frac{y_i - x_i^T\beta}{\sigma}\right), \tag{2.6}$$

where the first product is over the N_0 observations for which $y_i = 0$ and the second product is over the N_1 observations for which $y_i > 0$.

$$\log L \;=\; \sum_0 \log(1 - \Phi_i) + \sum_1 \log\left(\frac{1}{(2\pi\sigma^2)^{1/2}}\right) - \sum_1 \frac{1}{2\sigma^2}(y_i - x_i^T\beta)^2, \tag{2.7}$$

where the summation $\sum_0$ is over the N_0 observations for which $y_i = 0$, and the summation $\sum_1$ is over the N_1 observations for which $y_i > 0$.

We now write down the first and second derivatives of $\log L$ with respect to β and σ^2. We use the following facts:

$$\begin{aligned}\frac{\partial \Phi_i}{\partial \beta} &= \phi_i x_i \\ \frac{\partial \Phi_i}{\partial \sigma^2} &= -\frac{1}{2\sigma^2} x_i^T \beta \phi_i \\ \frac{\partial \phi_i}{\partial \beta} &= -\frac{1}{2\sigma^2} x_i^T \beta \phi_i x_i \\ \frac{\partial \phi_i}{\partial \sigma^2} &= \frac{(x_i^T \beta)^2 - \sigma^2}{2\sigma^4} \phi_i. \end{aligned} \tag{2.8}$$

Using these results, we get the first order conditions for a maximum as

$$\frac{\partial \log L}{\partial \beta} = -\sum_0 \frac{\phi_i x_i}{1 - \Phi_i} + \frac{1}{\sigma^2} \sum_1 (y_i - x_i^T \beta) x_i = 0 \tag{2.9}$$

$$\frac{\partial \log L}{\partial \sigma^2} = \frac{1}{2\sigma^2} \sum_0 \frac{x_i^T \beta \phi_i}{1 - \Phi_i} - \frac{N_1}{\sigma^2} + \frac{1}{2\sigma^4} \sum_1 (y_i - x_i^T \beta)^2 = 0. \tag{2.10}$$

Premultiplying (2.9) by $x_i^T \beta$ and adding the result to (2.10), we get

$$\sigma^2 = \frac{1}{N_1} \sum_1 (y_i - x_i^T \beta) y_i = \frac{Y_1^T (Y_1 - X_1 \beta)}{N_1}, \tag{2.11}$$

where was used the notations in (2.5).

Also, after multiplying throughout by σ^2, equation (2.9) can be written as

$$-X_0^T \gamma_0 + \frac{1}{\sigma} X_1^T (Y_1 - X_1 \beta) = 0, \tag{2.12}$$

or

$$\beta = (X_1^T X_1)^{-1} X_1^T Y_1 - \sigma(X_1^T X_1)\gamma_0 \tag{2.13}$$

$$= \hat{\beta}_{OLS} - \sigma(X_1^T X_1)\gamma_0, \tag{2.14}$$

where $\hat{\beta}_{OLS}$ is the ordinary least squares estimator for β obtained from the N_1 nonzero observations on the response $y = (y_1, \ldots, y_N)$.

Equation (2.14) shows the relationship between the maximum likelihood estimator for β and the ordinary least squares estimator obtained from the nonzero observations of y.

2.4 Some important results on Tobit models.

Fair (1977) suggested an iteration method for obtaining the maximum likelihood estimates of β and σ^2 using equation (2.14). The method he suggested is the following:

(1) Compute β_{LS}, and calculate $(X_1^T X_1)^{-1} X_0$.

(2) Choose a value of β, say $\beta^{(1)}$, and compute σ^2 from (2.11). If this value of σ^2 is less than or equal to zero, take for the value of σ^2 some small positive number. Let $\sigma^{(1)}$ denote the squared root of this chosen value of σ^2.

(3) Compute the vector γ_0 using $\beta^{(1)}$ and $\sigma^{(1)}$. Denote this by $\gamma_0^{(1)}$.

(4) Compute β from equation (2.14) using $\sigma^{(1)}$ and $\gamma_0^{(1)}$. Denote this value by $\tilde{\beta}^{(1)}$. Let

$$\beta^{(2)} = \beta^{(1)} + \lambda(\tilde{\beta}^{(1)} - \beta^{(1)}), \quad 0 < \lambda \leq 1.$$

(5) Using $\beta^{(2)}$, go to step (2), and repeat the process until the iterations converge.

Fair (1977) suggested using zero as the starting value for $\beta^{(1)}$ if there is a large number of zero observations and using the least squares estimator $\hat{\beta}_{OLS}$ as the starting value if there is a large number of nonzero observations. He also suggested using $\lambda = 0.4$ and an iteration limit of twenty as a good initial strategy (these values worked best in the examples he looked at).

The Newton-Raphson method uses the matrix of second derivatives. These ingredients are as follows:

$$\frac{\partial \log L}{\partial\beta\partial\beta^T} = -\sum_0 \frac{\phi_i}{(1-\Phi_i)^2}\left[\phi_i - \frac{1}{\sigma^2}(1-\Phi_i)x_i^T\beta\right]x_i x_i^T - \frac{1}{\sigma^2}\sum_1 x_i x_i^T \tag{2.15}$$

$$\frac{\partial^2 \log L}{\partial\sigma^2\partial\beta} = -\frac{1}{2\sigma^2}\sum_0 \frac{\phi_i}{(1-\Phi_i)^2}\left[\frac{1}{\sigma^2}(1-\Phi_i)(x_i^T\beta)^2 - (1-\Phi_i) - x_i^T\beta\phi_i\right]x_i - \frac{1}{\sigma^4}\sum_1 (y_i - x_i^T\beta)x_i \tag{2.16}$$

$$\frac{\partial^2 \log L}{\partial(\sigma^2)^2} = \frac{1}{4\sigma^4}\sum_0 \frac{\phi_i}{(1-\Phi_i)^2}\left[\frac{1}{\sigma^2}(1-\Phi_i)(x_i^T\beta)^3 - 3(1-\Phi_i)x_i^T\beta - (x_i^T\beta)^2\Phi_i\right] + \frac{N_1}{2\sigma^4} - \frac{1}{\sigma^6}\sum_1 (y_i - x_i^T\beta)^2. \tag{2.17}$$

In the method of scoring, one uses the probability limits of these second derivatives used in the Newton-Raphson method of iteration. This simplifies the expressions considerably. Using the first order conditions for the maximum given by equations (2.9) and (2.10) and substituing

$$\sum(1-\Phi_i)Z_i \quad \text{by} \quad \sum_0 Z_i \quad \text{and}$$
$$\sum \Phi_i Z_i \quad \text{by} \quad \sum_1 Z_i$$

where the summations now run over all the observations, the second derivatives can be written compactly (Amemiya, 1973, p. 1007) as

$$\frac{\partial^2 \log L}{\partial\beta\partial\beta^T} = -\sum a_i x_i x_i^T \quad (2.18)$$

$$\frac{\partial^2 \log L}{\partial\sigma^2\partial\beta} = -\sum b_i x_i \quad (2.19)$$

$$\frac{\partial^2 \log L}{\partial(\sigma^2)^2} = -\sum c_i. \quad (2.20)$$

The summations now run over all the N observations, and

$$a_i = -\frac{1}{\sigma^2}\left(Z_i\phi_i - \frac{\phi_i^2}{1-\Phi_i} - \Phi_i\right) \quad (2.21)$$

$$b_i = \frac{1}{2\sigma^3}\left(Z_i^2\phi_i + \phi_i - \frac{Z_i\phi_i^2}{1-\Phi_i}\right) \quad (2.22)$$

$$c_i = -\frac{1}{4\sigma^4}\left(Z_i^3\phi_i + Z_i\phi_i - \frac{Z_i^2\phi_i^2}{1-\Phi_i} - 2\Phi_i\right), \quad (2.23)$$

where

$$Z_i = \frac{x_i^T\beta}{\sigma}. \quad (2.24)$$

The asymptotic covariance matrix of the estimates of (β, σ^2) can be estimated as V^{-1}, where

$$V = \begin{pmatrix} \sum a_i x_i x_i^T & \sum b_i x_i \\ \sum b_i x_i^T & \sum c_i \end{pmatrix}, \quad (2.25)$$

and a_i, b_i, and c_i are as defined in (2.21)-(2.23).

2.4.1 A re-parametrization of the Tobit model.

In his study, Tobin (1958) reparametrized the model given by equation (2.1) by dividing throughout by σ. We can write the reparametrized model as

$$\tau y_i = \begin{cases} x_i^T\gamma + v_i, & \text{if } x_i^T\gamma + v_i > 0 \\ 0 & \text{otherwise} \end{cases}, \tag{2.26}$$

where $\tau = 1/\sigma$, $\gamma = \tau\beta$, and $v_i = u_i/\sigma$ has the standard normal distribution $N(0,1)$. The log-likelihood can now be written as

$$\log L = \sum_0 \log\Phi(-x_i^T\gamma) + N_1 \log h - \frac{1}{2}\sum_1 (\tau y_i - x_i^T\gamma)^2. \tag{2.27}$$

Let $\theta = (\gamma^T, \tau)^T$. Olsen (1978 b) showed that for the foregoing likelihood function, the matrix $\partial^2 \log L/\partial\theta\partial\theta^T$ is negative semidefinite.

For the reparametrized version, the expressions for the second derivatives are less cumbersome than the expressions in equations (2.15) through (2.17). The normal equations for the model are

$$\frac{\partial \log L}{\partial \gamma} = -\sum_0 \frac{\phi(-x_i^T\gamma)}{\Phi(-x_i^T)\gamma} x_i + \sum_1 (\tau y_i - x_i^T\gamma) x_i = 0 \tag{2.28}$$

$$\frac{\partial \log L}{\partial \tau} = \frac{N_1}{\tau} - \sum_1 (\tau y_i - x_i^T\gamma) y_i = 0. \tag{2.29}$$

The second derivatives are

$$\frac{\partial^2 \log L}{\partial \gamma \gamma^T} = \sum_0 \frac{\phi(-x_i^T\gamma)}{\Phi(-x_i^T\gamma)} \left(x_i^T\gamma - \frac{\phi(-x_i^T\gamma)}{\Phi(-x_i^T\gamma)} \right) x_i x_i^T - \sum_1 x_i x_i^T \tag{2.30}$$

$$\frac{\partial^2 \log L}{\partial \gamma \partial \tau} = \sum_1 y_i x_i \tag{2.31}$$

$$\frac{\partial^2 \log L}{\partial \tau^2} = -\frac{N_1}{\tau^2} - \sum_1 y_i^2. \tag{2.32}$$

To show that the matrix of second partial is negative semidefinite, we proceed as follows. Let

$$X = \begin{pmatrix} X_0 \\ X_1 \end{pmatrix}, \quad Y = \begin{pmatrix} Y_0 \\ Y_1 \end{pmatrix}, \quad \theta = \begin{pmatrix} \gamma \\ \tau \end{pmatrix},$$

where X_0, X_1, Y_0 and Y_1 are defined in (2.5). Then the matrix of second derivatives given by equations (2.30) and (2.31) can be written as

$$\frac{\partial^2 \log L}{\partial \theta \partial \theta^T} = -\begin{pmatrix} X^T \\ Y^T \end{pmatrix} \begin{pmatrix} D & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} X & Y \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & N_1/\tau^2 \end{pmatrix}, \tag{2.33}$$

where D is a diagonal matrix whose ith diagonal element is

$$D_i = -\frac{\phi(-x_i^T\gamma)}{\Phi(-x_i^T\gamma)}\left(x_i^T\gamma - \frac{\phi(x_i^T\gamma)}{\Phi(-x_i^T\gamma)}\right). \tag{2.34}$$

Thus, all that needs to be shown is that D_i is positive. For this, we note first that if w follows a standard normal distribution, and if we consider values of $w > c$, then, obviously,

$$E(w \mid w > c) > c.$$

Also, (see Appendix A)

$$E(w \mid w > 0) = \frac{\phi(c)}{1 - \Phi(c)}.$$

Hence,

$$c - \frac{\phi(c)}{1 - \Phi(c)} < 0. \tag{2.35}$$

Substituting $x_i^T\gamma$ for c, we see that D_i as given by equation (2.34) is positive. This proves the required result.

The implication of this result is that the likelihood function for the Tobit model has a single maximum. Thus, no matter what the starting value, as long as the iterative process is continued to obtain a solution, that solution will be the global maximum of the likelihood function. Also, as proved by Amemiya (1973), this estimator is consistent and asymptotically normal.

2.4.2 Two stage estimation of the Tobit model.

Amemiya (1973) criticized Tobin's initial estimator, saying that it was not consistent, and he suggested an initial consistent estimator. He also showed that the second round estimator that results from taking one iteration from this initial consistent estimator is asymptotically equivalent to the maximum likelihood estimator. Many empirical researchers found Tobin's initial estimator satisfactory, and Olsen's proof that the likelihood function for the Tobit model has a single maximum explains why they did not have trouble with their initial estimator. However, even though, the estimation method suggested by Amemiya is not useful for the Tobit model, it will be useful for more complicated models in which the likelihood function is likely to have multiple maxima. Hence, we shall review the method and some of its extensions here.

Considering the model given by equation (2.1) and the nonzero observations y_i we get

$$E(y_i \mid y_i > 0) = x_i^T\beta + E(u_i \mid u_i > -x_i^T\beta) = x_i^T\beta + \sigma\frac{\phi_i}{\Phi_i}. \qquad (2.36)$$

Also, using the formulas in the Appendix A for the second moments of the truncated normal and simplifying, we get

$$E(y_i^2 \mid y_i > 0) = x_i^T\beta E(y_i) + \sigma^2.$$

We can write this as

$$y_i^2 = x_i^T\beta y_i + \sigma^2 + \eta_i,$$

where $E(\eta_i) = 0$. However, this equation cannot be estimated by OLS, because $Cov(\eta_i, x_i y_i) \neq 0$. What Amemiya suggested is to regress y_i on x_i and higher powers of x_i, get $\hat{y}_i$ (the estimated value of y_i) from this equation, and use $x_i\hat{y}_i$ as instrumental variables to estimate this equation. Although it is theoretically appealing, the practical usefulness of this instrumental-variable approach is not known.

An alternative procedure first used by Heckman (1976 *b*) is given by the following. Because the likelihood function for the probit model is well behaved, we define a dummy variable

$$I_i = \begin{cases} 1 & \text{if } y_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

Then, using the probit model, we get consistent estimates of β/σ. Using these, we get estimated values of ϕ_i and Φ_i. Now we get consistent estimates of β and σ by estimating equation (2.36) by OLS, with $\hat{\phi}_i/\hat{\Phi}_i$ as the explanatory variable in place of ϕ_i/Φ_i.

Instead of using only the nonzero observations on y_i, if we use all the observations, we get

$$\begin{aligned} E(y_i) &= P(y_i > 0) \cdot E(y_i \mid y_i > 0) + P(y_i = 0) \cdot E(y_i \mid y_i = 0) \\ &= \Phi_i \left(x_i^T \beta + \sigma \frac{\phi_i}{\Phi_i} \right) + (1 - \Phi_i)0 \\ &= \Phi_i x_i^T \beta + \sigma \phi_i \qquad (2.37) \end{aligned}$$

Thus, after getting estimates of ϕ_i and Φ_i, we estimate equation (2.37) by OLS.

2.4.3 Prediction in the Tobit model.

The prediction of y_i, given x_i, can be obtained from the different expectations functions we have given. Note that if we define the model in (2.1) in terms of a latent variable framework, with

$$y_i^* \;=\; x_i^T\beta + u_i, \quad \text{and} \quad E(u_i) = 0,$$

modeling the latent variable y_i^* (say, desired or potential expenditures), and if we define y_i, the observed variable, as

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

then clearly $E(y_i^*) = x_i^T\beta$.

Thus, after estimating β , we can get predictions of the latent variable from this equation. There are, in addition, two other predictions we can make, and these are predictions about the observed y_i, given the information that it is greater than zero and not given any such information. These predictions are given by equations (2.36) and (2.37) . What (2.36) gives is $E(y_i^* \mid y_i^* > 0)$ or $E(y_i \mid y_i > 0)$, that is the mean of the positive $y's$. What (2.37) gives is $E(y_i)$, that is the mean of all observed $y's$, positive and zero. Note that this is not $E(y_i^*)$ which is the mean of potential $y's$. Thus, there are three predictions: one for the latent variable and two for the observed variable (with or without the information that the observed variable is greater than zero). Note that $E(y_i) = \Phi_i E(y_i \mid y_i > 0)$. Corresponding to these three expectations functions we have the following derivatives to predict the effects of changes in the exogenous variables.

Denote $z_i = x_i^T \beta/\sigma$ and β_j the jth component of β. We drop the subscript i, which refers to the ith observation. Then, using the formulas for derivatives in the Appendix, we have, after simplifications,

$$\begin{aligned} \frac{\partial E(y^*)}{\partial x_j} &= \beta_j \\ \frac{\partial E(y)}{\partial x_j} &= \Phi(z)\beta_j \end{aligned}$$

and

$$\frac{\partial E(y \mid y^* > 0)}{\partial x_j} = \beta_j \left(1 - z\frac{\phi(z)}{\Phi(z)} - \left(\frac{\phi(z)}{\Phi(z)} \right)^2 \right)$$

2.5 Asymptotic theory: consistency of the estimators.

From figure (2.1) it is clear that the least squares regression of expenditure on income using all the observations including zero expenditures yields biased estimates.

Although it is not so clear from the figure, the least square regression using only the positive expenditures also yields biased estimates. These facts can be mathematically demonstrated as follows.

First, consider the regression using only positive observations of y_i. Recalling equation (2.36), we obtain that the last term in the middle is generally nonzero (even without assuming that u_i is normal). This implies the biasedness of the OLS estimator using positive observation on y_i under more general models that the Tobit model.

Goldberger (1981) evaluated the asymptotic bias (the probability limit minus the true value) assuming that the elements of x_i (except the first element, which is assumed to be a constant) are normally distributed. More specifically, Goldberger rewrote the model in (2.1) as

$$y_i = \begin{cases} \beta_0 + \bar{x}_i^T \beta_1 + u_i & \text{if } \beta_0 + \bar{x}_i^T \beta_1 + u_i > 0 \\ 0 & \text{otherwise} \end{cases}, \tag{2.38}$$

and assumed $\bar{x}_i \sim N(0, \Sigma)$, distributed independently of u_i. (here the assumption of zero mean involves no loss of generality because a nonzero mean can be absorbed into β_0). Under this assumption he obtained

$$plim(\hat{\beta}_1) = \frac{1-\Gamma}{1-\rho^2\Gamma}\beta_1, \quad (2.39)$$

where

$$\Gamma = \sigma_y^{-1}\lambda(\beta_0/\sigma_y)\left(\beta_0 + \sigma_y\lambda(\beta_0/\sigma_y)\right)$$
$$\rho^2 = \sigma_y^{-2}\beta_1^T\Sigma\beta_1$$

where

$$\sigma_y^2 = \sigma^2 + \beta_1^T\Sigma\beta_1$$
$$\lambda(z) = \frac{\phi(z)}{\Phi(z)^2}.$$

It can be shown that $0 < \Gamma < 1$ and $0 < \rho^2 < 1$; therefore (2.39) shows that $\hat{\beta}_1$ shrinks β_1 toward 0. However, the result may not hold if $\bar{x}_i$ is not normal.

Consider the regression using all the observations of y_i, both positive an 0. To see that the least squares estimator is also biased in this case, we should look at the unconditional mean of y_i,

$$E(y_i) = x_i^T\beta\Phi_i + \sigma\phi_i. \quad (2.40)$$

Writing again (2.1) as (2.38) and using the same assumptions as Goldberger, Greene (1981) showed

$$plim(\tilde{\beta}_1) \quad = \quad \Phi(\beta_0/\sigma_y)\beta_1, \tag{2.41}$$

where $\tilde{\beta}_1$ is the LS estimator of β_1 in the regression of y_i on $\bar{x}_i$ using all the observations. This result is more useful than (2.39) because it implies than $(N_1/N)\tilde{\beta}_1$ is a consistent estimator estimator of β_1, where N_1 is the number of positive observations of y_i. A simple consistent estimator of β_0 can be similarly obtained. Greene (1983) gave the asymptotic variances of these estimators. However, we cannot use this estimator without knowing its properties when the true distribution of $\bar{x}_i$ is not normal.

2.6 Summary of the Tobit model.

A truncated dependent variable arises when values of the dependent variable are excluded from the sample, either by choice of the researcher to use a (non-randomly) selected sample of the population.

The aim of this chapter was to introduce those methodological issues about the Tobit model of major relevance, and the appropriate statistical methods developed to allow for consistent and efficient estimation of models that involve a limited dependent variable. The basic model assumes homoscedastic error variances but this assumption is easily relaxed to allow for a relatively general form of heteroscedasticity. Most damaging to the Tobit model is violation of the assumption of normality of u_i, since violation of this assumption produces inconsistent maximum likelihood estimates (see Greene, 2002, pp.771-772).

Under the reparametrized version of the Tobit model, Olsen's result (1978) has the implication that no matter what the starting value is, if the iterative process yields a solution it will be the global maximum of the likelihood function and hence such estimator will be consistent and asymptotically normal.

Chapter 3

The t-Censored Regression Model

"My brain grew clearer, I understood that I was close to total collapse. I put my hands against the wall and shoved to push myself away from it. The street was still dancing around. I began to hiccup from fury, and struggled with every bit of energy against my collapse, fought a really stout battle not to fall down. I didn't want to fall, I wanted to die standing".

Paul Auster, *The Art of Hunger*

3.1 Introduction.

An natural way of extending Tobin's model is to suppose that the distribution of the perturbations is not normal, allowing us to assume some other symmetric distribution like the Student-t. This selection is based on the robustness that the Student-t posses; the t distribution provides a useful extension of the normal distribution for the statistical modeling of data sets involving errors with tails heavier than those of the normal distribution (see e.g. Lange et al., 1989).

3.2 The Student-t censored regression model

In this section, we study the t-censored regression model which is defined by assuming in (2.1) that the disturbances $u_1, \ldots, u_N$ are iid $t(0, \sigma^2, \nu)$, where $t(\mu, \sigma^2, \nu)$ denotes the t-distribution with location parameter μ, scale parameter σ^2 and ν degrees of freedom. This is equivalent to consider that the unobserved random random variables $y_1^*, \ldots, y_N^*$ are independent, with $y_i^* \sim t(x_i^T\beta, \sigma^2, \nu)$, i.e. with pdf given by

$$f(y_i^*; \beta, \sigma^2) = \frac{1}{\sigma} c(\nu) \left\{ 1 + \frac{z_i^2}{\nu} \right\}^{-(\nu+1)/2},$$

where $z_i = (y_i^* - x_i^T\beta)/\sigma$ and $c(\nu) = \Gamma((\nu+1)/2)/\sqrt{\pi\nu}\,\Gamma(\nu/2)$. In that follows the standardized $t(0, 1, \nu)$-pdf will be denotes by $t(z; \nu) = c(\nu)\{1 + z^2/\nu\}^{-(\nu+1)/2}$ and its cdf by $T(z; \nu)$. Considering this notation, we have $P(y_i = 0) = 1 - T(x_i^T\beta/\sigma; \nu)$ for $i \in I_0$ and $y_i \sim t(x_i^T\beta, \sigma^2, \nu)$ for $i \in I_1$, where N_0 and N_1 are the number of observations on the sets $I_0 = \{i : y_i = 0\}$ and $I_1 = \{i : y_i > 0\}$, respectively. Hence, the t-censored regression likelihood will be given by

$$L(\beta, \sigma, \nu) = \prod_{i \in I_0} \left[1 - T\left(\frac{x_i^T\beta}{\sigma}; \nu \right) \right] \times \prod_{i \in I_1} \left[\frac{1}{\sigma} t\left(\frac{y_i - x_i^T\beta}{\sigma}; \nu \right) \right]. \quad (3.1)$$

Since (3.1) reduces to the Cauchy censored regression likelihood function when $\nu = 1$ and also it converges to the Tobit likelihood function in (2.6) as $\nu \to \infty$, the t-censored regression model provides a robust generalization of the Tobit model.

3.2.1 The mean and variance of a censored t-response

Writing the observed response $y = \max\{0, \mu + \sigma u\}$, where $\mu = x^T\beta$ and $u \sim t(0,1,\nu)$, we have for the mean of y that

$$E(y) = E(y|y>0)P(y>0) + E(y|y=0)P(y=0) = \{\mu + \sigma E(u|u>-c)\}T(c;\nu),$$

where $c = \mu/\sigma$. Similarly, but after some straightforward algebra, we obtain for the variance of y that

$$Var(y) = \{\mu^2 + 2\mu\sigma E(u|u>-c) + \sigma^2 E(u^2|u>-c)\}T(c;\nu)\{1 - T(c;\nu)\}.$$

Consequently, we need to compute the truncated moment $E(u^k|u>-c)$ for $k = 1,2$ in order to obtain the mean and variance of y. We give these results in the Appendix A5, for $k = 1,2,3,4$, following from there that:

$$\begin{aligned} E(u|u>-c) &= \left(\frac{\nu}{\nu-1}\right)\left(1+\frac{c^2}{\nu}\right)\frac{t(z;\nu)}{T(z;\nu)}, \quad \nu > 1, \\ E(u^2|u>-c) &= \left(\frac{\nu}{\nu-2}\right)\frac{T\left(c_{-2};\nu-2\right)}{T(c;\nu)} - cE(u|u>-c), \quad \nu > 2, \end{aligned}$$

where $c_{-2} = \sqrt{\frac{\nu-2}{\nu}}\,c$. We note for $\nu > 2$ that the first truncated mean can be rewrite also as $E(u|u>-c) = \sqrt{\frac{\nu}{\nu-2}}\,\frac{t(c_{-2};\nu-2)}{T(c;\nu)}$. Hence, we obtain for the mean and variance of the t-censored model that

$$E(y) = \mu T(c;\nu) + \sigma \left(\frac{\nu}{\nu-1}\right)\left(1+\frac{c^2}{\nu}\right) t(c;\nu) \tag{3.2}$$

and

$$\begin{aligned} Var(y) &= \{1 - T(c;\nu)\} \\ &\times \left\{\mu^2 T(c;\nu) + \mu\sigma\sqrt{\frac{\nu}{\nu-2}}\, t\left(c_{-2};\nu-2\right) + \sigma^2\left(\frac{\nu}{\nu-2}\right) T\left(c_{-2};\nu-2\right)\right\} \end{aligned} \tag{3.3}$$

By letting $\nu \to \infty$, we obtain from (3.2) and (3.3) the following results for the normal censored mean and variance, respectively:

$$E(y) \to \mu\Phi(c) + \sigma\phi(c) \quad \text{and} \quad Var(y) \to \{1 - \Phi(c)\}\left\{\mu^2\Phi(c) + \mu\sigma\phi\left(c\right) + \sigma^2\Phi\left(c\right)\right\}.$$

3.2.2 A two-stage estimation procedure

Following Heckman (1976) (see also Maddala, 1983, Section 6.5), an initial consistent estimator can be obtained by maximizing the log-likelihood function corresponding to the dummy variables

$$d_i = \begin{cases} 1 & \text{if } y_i > 0 \\ 0 & \text{if } y_i = 0, \end{cases}$$

which is given by

$$\sum_{i\in I_0}(1-d_i)\log\{1 - T(c_i;\nu)\} + \sum_{i\in I_1} d_i \log T(c_i;\nu), \quad \text{with} \quad c_i = x_i^T\beta/\sigma,$$

and is well-behaved (see e.g. Amemiya, 1984). Then, we get consistent estimates of β/σ and ν. Using these estimates, we get estimated values of c_i, $t_i = t(c_i;\nu)$ and $T_i = T(c_i;\nu)$. Now, considering equation (3.2) we have for $\nu > 1$ that

$$E(y_i) = (T_i x_i)^T \beta + (c_i^2 t_i)\sigma_\nu + t_i \nu_\sigma, \quad i = 1, \ldots, N, \tag{3.4}$$

where $\sigma_\nu = \sigma/(\nu-1)$ and $\nu_\sigma = \nu\sigma/(\nu-1)$. Thus, after getting estimates $\tilde{c}_i$, $\tilde{t}_i$ and $\tilde{T}_i$ of c_i, t_i and T_i, respectively, we estimate equation (3.4) by OLS. This yields the following OLS estimates of β, σ_ν and ν_σ:

$$\begin{pmatrix} \tilde{\beta} \\ \tilde{\sigma}_\nu \\ \tilde{\nu}_\sigma \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^N \tilde{T}_i^2 x_i x_i^T & \sum_{i=1}^N \tilde{c}_i^2 \tilde{t}_i \tilde{T}_i x_i & \sum_{i=1}^N \tilde{t}_i \tilde{T}_i x_i \\ \sum_{i=1}^N \tilde{c}_i^2 \tilde{t}_i \tilde{T}_i x_i^T & \sum_{i=1}^N \tilde{c}_i^4 \tilde{t}_i^2 & \sum_{i=1}^N \tilde{c}_i^2 \tilde{t}_i^2 \\ \sum_{i=1}^N \tilde{t}_i \tilde{T}_i x_i & \sum_{i=1}^N \tilde{c}_i^2 \tilde{t}_i^2 & \sum_{i=1}^N \tilde{t}_i^4 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^N \tilde{T}_i x_i y_i \\ \sum_{i=1}^N \tilde{c}_i^2 \tilde{t}_i y_i \\ \sum_{i=1}^N \tilde{t}_i y_i \end{pmatrix}.$$

Finally, from these results we can recuperate the estimates for σ and ν given by $\tilde{\sigma} = (\tilde{\nu}-1)\tilde{\sigma}_\nu$ and $\tilde{\nu} = \tilde{\nu}_\sigma/\tilde{\sigma}_\nu$.

3.2.3 Maximum likelihood estimation

By convenience, we consider in that follows the Olsen's (1978) re-parametrization $\gamma = \tau\beta$ and $\tau = 1/\sigma$. Under this new parametrization, the log-likelihood function for $\theta = (\gamma^T, \tau, \nu)^T$ obtained from (3.1) is given by:

$$\log L(\theta) = \sum_{i \in I_0} \log\left\{1 - T\left(x_i^T \gamma; \nu\right)\right\} + N_1 \log \tau + \sum_{i \in I_1} \log t\left(z_i; \nu\right), \tag{3.5}$$

where $z_i = \tau y_i - x_i^T \gamma$, and

$$\log t(z_i;\nu) = \log \Gamma\left(\frac{\nu+1}{2}\right) - \log \Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\log(\pi\nu) - \frac{\nu+1}{2}\log\left(1 + \frac{z_i^2}{\nu}\right). \tag{3.6}$$

To derive the scores components $S_\gamma = (\partial \log L/\partial\gamma)$, $S_\tau = (\partial \log L/\partial\tau)$ and $S_\nu = (\partial \log L/\partial\nu)$, consider first the following partial derivatives:

$$
\begin{aligned}
\frac{\partial z_i}{\partial \gamma} &= -x_i, \quad \frac{\partial z_i}{\partial \tau} = y_i, \quad \frac{\partial T\left(x_i^T\gamma;\nu\right)}{\partial\gamma} = t\left(x_i^T\gamma;\nu\right)x_i \\
\frac{\partial \log t(z;\nu)}{\partial\nu} &= \frac{1}{2}\left\{\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - \frac{1}{\nu}\right\} - \frac{1}{2}\log\left(1+\frac{z_i^2}{\nu}\right) \\
&\quad + \frac{1}{2}\left(\frac{\nu+1}{\nu}\right)\frac{z_i^2}{\nu}\left(1+\frac{z_i^2}{\nu}\right)^{-1} \\
\frac{\partial T\left(x_i^T\gamma;\nu\right)}{\partial\nu} &= \frac{1}{2}\left\{\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - \frac{1}{\nu}\right\} T\left(x_i^T\gamma;\nu\right) - \frac{1}{2}b_{01}\left(x_i^T\gamma;\nu\right)T\left(x_i^T\gamma;\nu\right) \\
&+ \frac{1}{2}\frac{1}{\nu}\{1-(x_i^T\gamma)r(x_i^T\gamma;\nu)\}T\left(x_i^T\gamma;\nu\right),
\end{aligned}
$$

where $\psi(x)$ is the digamma function, and $b_{km}(c_k;\nu+k)$ is the truncated moment defined by

$$
b_{km}(c_k;\nu) = \int_{-\infty}^{c_k} z^k\left\{\log\left(1+\frac{z^2}{\nu}\right)\right\}^m \frac{t(z;\nu+k)}{T(c_k;\nu+k)}dz, \quad \text{with} \quad c_k = \sqrt{\frac{\nu}{\nu+k}}\, c. \tag{3.7}
$$

A proof of the above partial derivative of $T(z;\nu)$ with respect to ν is given in the Appendix B. Thus, by considering also the ratios

$$
R(x_i^T\gamma;\nu) = \frac{T(x_i^T\gamma;\nu)}{1-T(x_i^T\gamma;\nu)} \quad \text{and} \quad r(-x_i^T\gamma;\nu) = \frac{t(-x_i^T\gamma;\nu)}{T(-x_i^T\gamma;\nu)} = \frac{t(x_i^T\gamma;\nu)}{1-T(x_i^T\gamma;\nu)},
$$

we obtain from (3.5) and (3.6) the following scores functions:

$$S_\gamma = -\sum_{i\in I_0} r(-x_i^T\gamma;\nu)x_i + (\nu+1)\sum_{i\in I_1}\frac{z_i}{\nu}\left(1+\frac{z_i^2}{\nu}\right)^{-1} x_i \tag{3.8}$$

$$S_\tau = \frac{N_1}{\tau} - (\nu+1)\sum_{i\in I_1}\frac{z_i}{\nu}\left(1+\frac{z_i^2}{\nu}\right)^{-1} y_i \tag{3.9}$$

$$\begin{aligned} S_\nu &= \frac{1}{2}\left\{\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - \frac{1}{\nu}\right\}\left\{N_1 - \sum_{i\in I_0} R(x_i^T\gamma;\nu)\right\} \\ &+ \frac{1}{2}\sum_{i\in I_0} R(x_i^T\gamma;\nu)b_{01}(x_i^T\gamma;\nu) - \frac{1}{2}\frac{1}{\nu}\sum_{i\in I_0}\{1-(x_i^T\gamma)r_i(x_i^T\gamma;\nu)\}R(x_i^T\gamma;\nu) \\ &- \frac{1}{2}\sum_{i\in I_1}\log\left(1+\frac{z_i^2}{\nu}\right) + \frac{1}{2}\left(\frac{\nu+1}{\nu}\right)\sum_{i\in I_1}\frac{z_i^2}{\nu}\left(1+\frac{z_i^2}{\nu}\right)^{-1}, \end{aligned} \tag{3.10}$$

where as was defined above $z_i = \tau y_i - x_i^T\gamma$.

Note when $\nu \to \infty$ that $r(z_i;\nu) \to r(z_i) = \frac{\phi(z_i)}{\Phi(z_i)}$ and $R(z_i;\nu) \to R(z_i) = \frac{\Phi(z_i)}{1-\Phi(z_i)}$, so that in equation (3.10) we have $S_\nu \to 0$, while the equations (3.8)-(3.9) reduce to the following Tobit's score functions:

$$S_\gamma = -\sum_{i\in I_0} r(-x_i^T\gamma)x_i + \sum_{i\in I_1}(\tau y_i - x_i^T\gamma)x_i \tag{3.11}$$

$$S_\tau = \frac{N_1}{\tau} - \sum_{i\in I_1}(\tau y_i - x_i^T\gamma)y_i. \tag{3.12}$$

Equaling to zero and applying some algebra, the (3.11) and (3.12) expressions can be reformulated in terms of the original parametrization $(\beta^T,\sigma)^T$ through the Tobit's likelihood equations given by (2.9) and (2.10).

3.3 The information matrix

In this section we obtain the observed and expected information matrix for the Student-t censored regression model from the second partial derivatives of the likelihood function (3.5). For this, we denote by $S_{\alpha\lambda} = \partial^2 \log L(\theta)/\partial\alpha\partial\lambda^T = \partial S_\alpha/\partial\lambda^T$ the second partial derivative with respect to the components α and λ of $\theta = (\gamma^T, \tau, \nu)^T$. Thus, from (3.8)-(3.9) we obtain after some extensive algebra (see Appendix B) that

$$
\begin{aligned}
S_{\gamma\gamma} &= -\sum_{i\in I_0} R(x_i^T\gamma;\nu) r(x_i^T\gamma;\nu) \left\{ R(x_i^T\gamma;\nu) r(x_i^T\gamma;\nu) \right. \\
&- \left. \frac{T(x_i^T\gamma_2;\nu+2)}{T(x_i^T\gamma;\nu)} (x_i^T\gamma_2) r\left(x_i^T\gamma_2;\nu+2\right) \right\} x_i x_i^T \\
&- \left(\frac{\nu+1}{\nu}\right) \sum_{i\in I_1} \left\{ \left(1+\frac{z_i^2}{\nu}\right)^{-1} - \frac{2}{\nu} z_i^2 \left(1+\frac{z_i^2}{\nu}\right)^{-2} \right\} x_i x_i^T \\
S_{\gamma\tau} &= \left(\frac{\nu+1}{\nu}\right) \sum_{i\in I_1} \left\{ \left(1+\frac{z_i^2}{\nu}\right)^{-1} - \frac{2}{\nu} z_i^2 \left(1+\frac{z_i^2}{\nu}\right)^{-2} \right\} y_i x_i \\
S_{\gamma\nu} &= -\frac{1}{2} \sum_{i\in I_0} \{R(x_i^T\gamma;\nu)\}^2 r(x_i^T\gamma;\nu) \left\{ \psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) \right. \\
&- \left. b_{01}\left(x_i^T\gamma;\nu\right) - \frac{1}{\nu}(x_i^T\gamma) r(x_i^T\gamma;\nu) \right\} x_i - \frac{1}{2} \sum_{i\in I_0} R(x_i^T\gamma;\nu) r(x_i^T\gamma;\nu) \\
&\times \left\{ \psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - \frac{1}{\nu} - \log\left(1+\frac{(x_i^T\gamma)^2}{\nu}\right) \right\} x_i \\
&- \frac{1}{2}\frac{1}{\nu}\left(\frac{\nu}{\nu+2}\right)^{1/2} \sum_{i\in I_0} \{1+R(x_i^T\gamma;\nu)\} T(x_i^T\gamma_2;\nu+2)(x_i^T\gamma_2)^2 r(x_i^T\gamma_2;\nu) x_i \\
&- \frac{1}{\nu^2} \sum_{i\in I_1} \left\{ z_i \left(1+\frac{z_i^2}{\nu}\right)^{-1} - \left(\frac{\nu+1}{\nu}\right) z_i^3 \left(1+\frac{z_i^2}{\nu}\right)^{-2} \right\} x_i
\end{aligned}
$$

$$
\begin{aligned}
S_{\tau\tau} &= -\frac{N_1}{\tau^2} - \left(\frac{\nu+1}{\nu}\right)\sum_{i\in I_1}\left\{\left(1+\frac{z_i^2}{\nu}\right)^{-1} - \frac{2}{\nu}z_i^2\left(1+\frac{z_i^2}{\nu}\right)^{-2}\right\}y_i^2\\
S_{\tau\nu} &= \frac{1}{\nu^2}\sum_{i\in I_1}\left\{z_i\left(1+\frac{z_i^2}{\nu}\right)^{-1} - \left(\frac{\nu+1}{\nu}\right)z_i^3\left(1+\frac{z_i^2}{\nu}\right)^{-2}\right\}y_i,
\end{aligned}
$$

and

$$
\begin{aligned}
S_{\nu\nu} &= \frac{1}{4}N_1\left\{\psi'\left(\frac{\nu+1}{2}\right) - \psi'\left(\frac{\nu}{2}\right) + \frac{2}{\nu^2}\right\} - \frac{1}{4}\sum_{i\in I_0}R\left(x_i^T\gamma;\nu\right)\{1+R(x_i^T\gamma;\nu)\}\\
&\times \left\{\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - b_{01}\left(x_i^T\gamma;\nu\right) - \frac{1}{\nu}(x_i^T\gamma)r(x_i^T\gamma;\nu)\right\}^2\\
&- \frac{1}{4}\sum_{i\in I_0}R(x_i^T\gamma;\nu)\left\{\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - b_{01}\left(x_i^T\gamma;\nu\right) - \frac{1}{\nu}(x_i^T\gamma)r(x_i^T\gamma;\nu)\right\}\\
&\times \left\{b_{01}(x_i^T\gamma;\nu) + \frac{1}{\nu}(x_i^T\gamma)r(x_i^T\gamma;\nu)\right\} - \frac{1}{4}\sum_{i\in I_0}R(x_i^T\gamma;\nu)\left\{\psi'\left(\frac{\nu+1}{2}\right) - \psi'\left(\frac{\nu}{2}\right)\right.\\
&- \frac{1}{\nu}(x_i^T\gamma)r(x_i^T\gamma;\nu)\left\{\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - \frac{3}{\nu} - \log\left(1+\frac{(x_i^T\gamma)^2}{\nu}\right)\right\}\\
&- \left.\frac{1}{\nu(\nu+2)}\{1+R(x_i^T\gamma;\nu)\}T(x_i^T\gamma_2;\nu+2)(x_i^T\gamma_2)^3 r(x_i^T\gamma_2;\nu+2)\right\}\\
&- \frac{1}{2}\frac{1}{\nu^3}\sum_{i\in I_1}\left\{2z_i^2\left(1+\frac{z_i^2}{\nu}\right)^{-1} - \left(\frac{\nu+1}{\nu}\right)z_i^4\left(1+\frac{z_i^2}{\nu}\right)^{-2}\right\},
\end{aligned}
$$

where $\gamma_k = \sqrt{\frac{\nu+k}{\nu}}\,\gamma$.

Using again that $r(z;\nu)\to r(z) = \frac{\phi(z)}{\Phi(z)}$, and $R(z;\nu)\to R(z) = \frac{\Phi(z)}{1-\Phi(z)}$, as $\nu\to\infty$, we obtain for the second partial derivatives that:

$$
\begin{aligned}
S_{\gamma\gamma} &\rightarrow -\sum_{i\in I_0} r(z)(x_i^T\gamma - r(z))x_ix_i^T - \sum_{i\in I_1} x_ix_i^T \\
S_{\gamma\tau} &\rightarrow \sum_{i\in I_1} y_ix_i \\
S_{\gamma\nu} &\rightarrow 0 \\
S_{\tau\tau} &\rightarrow -\frac{N_1}{\tau^2} - \sum_{i\in I_1} y_i^2 \\
S_{\tau\nu} &\rightarrow 0 \\
S_{\nu\nu} &\rightarrow 0,
\end{aligned}
$$

which agree with the corresponding Tobit results given in (2.30)-(2.32)

3.3.1 Expected information matrix.

To obtain the expected information matrix $I(\theta) = -E\{\partial^2 \log L(\theta)/\partial\theta\partial\theta^T\}$, we need truncated expectations the form of $E\left\{z^q\left(1+\frac{z^2}{\nu}\right)^{-s}|z<c\right\}$, where $z \sim t(0,1,\nu)$, which are obtained from the results in the Appendix A. We need also truncated expectations defined by (3.7) which need to be computed numerically. Hence, using appropriately those results, we obtain that $I(\theta)$ has (block-matrix) elements $I_{\alpha\lambda} = -E\{S_{\alpha\lambda}\}$ given by

$$
\begin{aligned}
I_{\gamma\gamma} &= \sum_{i \in I_0} R(x_i^T\gamma;\nu) r(x_i^T\gamma;\nu) \left\{ R(x_i^T\gamma;\nu) r(x_i^T\gamma;\nu) \right.\\
&- \left. \frac{T(x_i^T\gamma_2;\nu+2)}{T(x_i^T\gamma;\nu)} (x_i^T\gamma_2) r\left(x_i^T\gamma_2;\nu+2\right) \right\} x_i x_i^T \\
&+ \left(\frac{\nu+1}{\nu+3}\right) \sum_{i \in I_1} \frac{T(x_i^T\gamma_2;\nu+2)}{T(x_i^T\gamma;\nu)} \left\{ 1 + \frac{2}{\nu+1} (x_i^T\gamma_2)\, r(x_i^T\gamma_2;\nu+2) \right\} x_i x_i^T \\
I_{\gamma\tau} &= -\frac{1}{\tau}\left(\frac{\nu-1}{\nu+3}\right) \sum_{i \in I_1} r(x_i^T\gamma;\nu) x_i - \frac{1}{\tau}\left(\frac{\nu+1}{\nu+3}\right) \sum_{i \in I_1} \frac{T(x_i^T\gamma_2;\nu+2)}{T(x_i^T\gamma;\nu)} (x_i^T\gamma) x_i \\
I_{\gamma\nu} &= \frac{1}{2} \sum_{i \in I_0} \{R(x_i^T\gamma;\nu)\}^2 r(x_i^T\gamma;\nu) \left\{ \psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - b_{01}\left(x_i^T\gamma;\nu\right) - \frac{1}{\nu}(x_i^T\gamma) r(x_i^T\gamma;\nu) \right\} x_i \\
&+ \frac{1}{2} \sum_{i \in I_0} R(x_i^T\gamma;\nu) r(x_i^T\gamma;\nu) \left\{ \psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - \frac{1}{\nu} - \log\left(1 + \frac{(x_i^T\gamma)^2}{\nu}\right) \right\} x_i \\
&+ \frac{1}{2}\frac{1}{\nu}\left(\frac{\nu}{\nu+2}\right)^{1/2} \sum_{i \in I_0} \{1 + R(x_i^T\gamma;\nu)\} T(x_i^T\gamma_2;\nu+2)(x_i^T\gamma_2)^2 r(x_i^T\gamma_2;\nu) x_i \\
&- \frac{1}{\nu(\nu+3)} \sum_{i \in I_1} \left\{ \left(\frac{\nu-1}{\nu+1}\right) r(x_i^T\gamma;\nu) + \left(\frac{\nu}{\nu+2}\right)^{1/2} \frac{T(x_i^T\gamma_2;\nu+2)}{T(x_i^T\gamma;\nu)} (x_i^T\gamma_2)^2\, r(x_i^T\gamma_2;\nu+2) \right\} x_i \\
I_{\tau\tau} &= \frac{2}{\tau^2}\left(\frac{\nu}{\nu+3}\right) N_1 + \frac{1}{\tau^2}\left(\frac{\nu+7}{\nu+3}\right) \sum_{i \in I_1} (x_i^T\gamma) r(x_i^T\gamma;\nu) + \frac{1}{\tau^2}\left(\frac{\nu+1}{\nu+3}\right) \sum_{i \in I_1} \frac{T(x_i^T\gamma_2;\nu+2)}{T(x_i^T\gamma;\nu)} (x_i^T\gamma)^2 \\
I_{\tau\nu} &= \frac{2}{\tau} \frac{1}{(\nu+1)(\nu+3)} N_1 + \frac{1}{\tau} \frac{4}{\nu(\nu+3)} \sum_{i \in I_1} (x_i^T\gamma)\, r(x_i^T\gamma;\nu),
\end{aligned}
$$

and

$$\begin{aligned}
I_{\nu\nu} &= -\frac{1}{2}\frac{1}{\nu^2}\frac{\nu^2+3\nu+6}{(\nu+1)(\nu+3)}N_1 - \frac{1}{4}\left\{N_1 - \sum_{i\in I_0} R(x_i^T\gamma;\nu)\right\}\left\{\psi'\left(\frac{\nu+1}{2}\right) - \psi'\left(\frac{\nu}{2}\right)\right\} \\
&+ \frac{1}{4}\sum_{i\in I_0}\{R(x_i^T\gamma;\nu)\}^2\left\{\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - b_{01}\left(x_i^T\gamma;\nu\right) - \frac{1}{\nu}(x_i^T\gamma)r(x_i^T\gamma;\nu)\right\}^2 \\
&+ \frac{1}{4}\left\{\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right\}\sum_{i\in I_0} R(x_i^T\gamma;\nu)\left\{\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - b_{01}\left(x_i^T\gamma;\nu\right)\right. \\
&- \left.\frac{2}{\nu}(x_i^T\gamma)r(x_i^T\gamma;\nu)\right\} + \frac{1}{4}\frac{1}{\nu}\sum_{i\in I_0} R(x_i^T\gamma;\nu)(x_i^T\gamma)r(x_i^T\gamma;\nu)\left\{\frac{3}{\nu} + \log\left(1 + \frac{(x_i^T\gamma)^2}{\nu}\right)\right\} \\
&- \frac{1}{4}\frac{1}{\nu(\nu+2)}\sum_{i\in I_0} R(x_i^T\gamma;\nu)\{1 + R(x_i^T\gamma;\nu)\}T(x_i^T\gamma_2;\nu+2)(x_i^T\gamma_2)^3 r(x_i^T\gamma_2;\nu+2) \\
&+ \frac{1}{2}\frac{1}{\nu(\nu+1)(\nu+3)}\sum_{i\in I_1}\left\{(x_i^T\gamma)r(x_i^T\gamma;\nu) + \left(\frac{\nu+1}{\nu+2}\right)\frac{T(x_i^T\gamma_2;\nu+2)}{T(x_i^T\gamma;\nu)}(x_i^T\gamma_2)^3\, r(x_i^T\gamma_2;\nu+2)\right\}.
\end{aligned}$$

It follows as $\nu \to \infty$ that

$$I(\theta) \to \begin{pmatrix} \sum_{i\in I_0} c_i x_i x_i^T + \sum_{i\in I_1} x_i x_i^T & -\frac{1}{\tau}\sum_{i\in I_1} d_i x_i & 0 \\ -\frac{1}{\tau}\sum_{i\in I_1} b_i x_i & \frac{2}{\tau^2}N_1 + \frac{1}{\tau^2}\sum_{i\in I_1}(x_i^T\gamma)b_i & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where $c_i = R(x_i^T\gamma)r(x_i^T\gamma)\left\{R(x_i^T\gamma)r(x_i^T\gamma) - (x_i^T\gamma_2)r\left(x_i^T\gamma_2\right)\right\}$ and $d_i = \left\{r(x_i^T\gamma;\nu) + (x_i^T\gamma)\right\}$.

Finally, for the expected information matrix of the original parametrization $\psi = (\beta, \sigma, \nu)$, we have $J(\psi) = (\partial\theta/\partial\psi)^T I(\theta)(\partial\theta/\partial\psi)$, where the Jacobian matrix $(\partial\theta/\partial\psi)$ is given by

$$\nabla = \begin{pmatrix} \frac{1}{\sigma}I_k & -\frac{1}{\sigma}\beta & 0 \\ 0 & -\frac{1}{\sigma^2} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

3.4 Consistency and asymptotic distributions.

Since our censored t-regression model satisfies the standard regularity conditions, we have that the maximum likelihood estimator $\hat{\theta} = (\hat{\gamma}^T, \hat{\tau}, \hat{\nu})$, of $\theta = (\gamma^T, \tau, \nu)^T$, verifies that

$$\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{\mathfrak{L}} N_{k+2}(0, I(\theta)^{-1}),$$

as $N \to \infty$, where $I(\theta)$ is the information matrix described above in Section 3.3.

Moreover, considering that the Jacobian of the transformation $(\gamma^T, \tau, \nu)^T \mapsto (\beta^T, \sigma, \nu)^T$ is given by $\nabla =$, previously defined, we have for the MLE $\hat{\eta} = (\hat{\beta}^T, \hat{\tau}, \hat{\nu})^T$ of $\eta = (\beta^T, \tau, \nu)^T$, that:

$$\sqrt{N}(\hat{\eta} - \eta) \xrightarrow{\mathfrak{L}} N_{k+2}(0, (\nabla I(\theta) \nabla^T)^{-1}).$$

Chapter 4

Applications and examples

"Memory, then, is not so much as the past contained within us, but as proof of our life in the present. If a man is to be truly present among his surroundings, he must be thinking not of himself, but of what he sees. He must forget himself in order to be there. And from that forgetfulness arises the power of memory. It is a way of living one's life so that nothing is ever lost ".

Paul Auster, *The Invention of Solitude*

4.1 A first illustrative example: Tobin's data

In this section we provide to the reader two examples were the Student-t model is applied. In the first place, we use the data used in Tobin's paper and included in the `survival` package from `R`. Two models are adjusted to the data set: the normal model, meaning, the Tobit model, and then with several values for the degrees of freedom parameter (known), we fit a independent t censored regression model. Thus, when $\nu \to \infty$, the behavior of the t-model should behave as in the normal case. For the data corresponding to Tobin's paper, the response is the **durable goods purchase** (the limited dependent variable), and the explanatory variables are **age** in years of the head of the spending unit and **liquidity ratio assets** (x 1000).

In both cases, the normal and the Student-t cases, the estimations were obtained throw a maximum likelihood numerical procedure. The standard error were taken as the solution of the expected information matrix, on each corresponding set of parameters. As can be seen in Table 4.2, the approximations of the estimates for the beta parameters are quite reasonable, and according to what we supposed, when ν tends to infinity, the normal model is achieved. The results of the fitted models are reported in the following tables:

Parameter	Estimate	SE	Z	p-value	log $L(\theta)$
Intercept	15.144	16.079	0.942	3.46e-01	-28.9
Age	-0.129	0.218	-0.590	5.55e-01	
Quant	-0.045	0.058	-0.782	4.34e-01	
log(σ)	1.717	0.310			

Table 4.1: *Normal disturbances case.*

	Parameter	Estimate	SE	Z	p-value	$\log L(\theta)$
$\nu = 7$	Intercept	11.712	16.291	0.719	4.72e-01	-29.2
	Age	-0.124	0.196	-0.633	5.27e-01	
	Quant	-0.0314	0.06	-0.519	6.04e-01	
	$\log(\sigma)$	1.609	0.340			
$\nu = 15$	Intercept	13.513	16.224	0.833	4.05e-01	-29.1
	Age	-0.127	0.208	-0.610	5.42e-01	
	Quant	-0.038	0.059	-0.651	5.15e-01	
	$\log(\sigma)$	1.667	0.323			
$\nu = 50$	Intercept	14.651	16.130	0.908	3.64e-01	-29
	Age	-0.128	0.215	-0.596	5.51e-01	
	Quant	-0.043	0.058	-0.741	4.59e-01	
	$\log(\sigma)$	1.702	0.314			
$\nu = \infty$	Intercept	15.139	16.080	0.942	3.46e-01	-28.9
	Age	-0.129	0.218	-0.590	5.55e-01	
	Quant	-0.0455	0.058	-0.781	4.35e-01	
	$\log(\sigma)$	1.717	0.310			

Table 4.2: *Independent Student-t case, ν parameter is considered as known.* $\log(\sigma)$ *is the natural logarithm of σ parameter.*

4.2 A Monte Carlo simulated data application

A Monte Carlo study is employed to check the behavior of our estimations. We are going to take the following configuration for the response variable: $Y = \max\{m(X) - e, 0\}, m(X) = 1 + 3X$, with $X \sim Uniform[0,1]$ and $e \sim t(0,1,\nu)$. The sample size is $N = 100, 250$ and 500. On the other hand, the number of Monte Carlo simulations is 1000. Several values for the ν parameter has been considered. The reported estimations for β_0, β_1, σ and ν are detailed in Table 4.3. The values for the beta parameters were $\beta_0 = 1, \beta_1 = 3$, and $\sigma = 1$. In all the cases, the estimated parameters were very closed to the real values. The estimation for the degrees of freedom parameter shows to be quite similar to the real one chosen. A Newton type optimization procedure was use to fit the simulated data in all cases. Standard errors are given in the parenthesis form, and were estimated from the expected information matrix. As the real value for ν grows, the estimation has a larger standard error, as in the $\nu = 12$ case, and probably many times over and under estimates the degrees of freedom significantly. In general, estimations has a good behavior under different values for the N sample size.

N		100	250	500	N		100	250	500
	ν	1.6682	1.5522	1.5232		ν	3.2345	3.0980	3.1298
		(0.7506)	(0.2919)	(0.1836)			(1.0621)	(0.7956)	(0.6569)
	β_0	1.0080	0.9974	0.9994		β_0	1.0535	1.0174	0.9982
		(0.2943)	(0.1745)	(0.1264)			(0.2184)	(0.1737)	(0.1103)
$\nu = 1.5$	β_1	2.9841	3.0002	3.0050	$\nu = 3$	β_1	2.9330	2.9491	3.0053
		(0.5164)	(0.3024)	(0.1350)			(0.4571)	(0.2931)	(0.1873)
	σ	0.9945	1.0042	1.0004		σ	1.0014	0.9770	0.9983
		(0.1685)	(0.0997)	(0.0697)			(0.1366)	(0.0839)	(0.0637)
	ν	6.0386	6.5105	6.5596		ν	12.1033	13.0817	12.9580
		(2.9812)	(4.0862)	(1.6359)			(6.8763)	(6.7403)	(6.3284)
	β_0	1.1317	0.9774	1.0313		β_0	1.3310	1.0289	1.0254
		(0.2111)	(0.1404)	(0.1098)			(0.1125)	(0.0876)	(0.2332)
$\nu = 6$	β_1	2.8511	3.0301	2.9069	$\nu = 12$	β_1	2.9234	3.0140	3.1071
		(0.4343)	(0.2115)	(0.1393)			(0.3608)	(0.1936)	(0.4273)
	σ	0.9803	0.9654	1.0073		σ	0.9763	1.0195	0.9982
		(0.1487)	(0.1063)	(0.0773)			(0.0260)	(0.0598)	(0.0170)

Table 4.3: *Independent Student-t case, ν unknown. All the parameters were estimated using a 1000 Monte Carlo sample size.*

Another simulation analysis was performed, considering a similar configuration for the response variable on the previous example: $Y = \max\{m(X) - e, 0\}, m(X) = 1+3X$, with $X \sim Uniform[0,1]$ and $e \sim (1-\varepsilon) \times N(0, \sigma^2) + \varepsilon \times N(0, 2\times\sigma^2)$, $\sigma = 3$, and $\varepsilon = 0.09999$(*contaminated* normal distribution). The sample size is $N = 250$.

The results are presented in the following tables:

Parameter	Estimate	SE	Z	p-value	$\log L(\theta)$
β_0	0.919	0.3843	2.39	1.68e-02	-549.3
β_1	2.962	0.6642	4.46	8.19e-06	
$\log(\sigma)$	1.061	0.0517			

Table 4.4: *Simulated analysis from ε-contaminated normal distribution data, fitted with a Student-t;* $\log(\sigma)$ *is the natural logarithm of σ parameter.*

	Parameter	Estimate	SE	Z	p-value	$\log L(\theta)$
$\boldsymbol{\nu = 7}$	β_0	0.917	0.3711	2.47	1.35e-02	-550.1
	β_1	2.862	0.6524	4.39	1.15e-05	
	$\log(\sigma)$	0.956	0.0583			
$\boldsymbol{\nu = 15}$	β_0	0.914	0.3769	2.42	1.53e-02	-549.3
	β_1	2.909	0.6572	4.43	9.56e-06	
	$\log(\sigma)$	1.007	0.0551			
$\boldsymbol{\nu = 50}$	β_0	0.916	0.3817	2.40	1.64e-02	-549.2
	β_1	2.945	0.6617	4.45	8.53e-06	
	$\log(\sigma)$	1.043	0.0528			
$\boldsymbol{\nu = \infty}$	β_0	0.918	0.384	2.39	1.67e-02	-549.2
	β_1	2.958	0.664	4.46	8.27e-06	
	$\log(\sigma)$	1.056	0.052			

Table 4.5: *Simulated analysis from ε-contaminated normal distribution data, fitted with a Student-t; he ν parameter is considered as known.* $\log(\sigma)$ *is the natural logarithm of σ parameter.*

The fitted values seems to behave quite similar to the normal case, when $\nu \to \infty$, which shows the capability of the t-model to fit the normal contaminated data.

Chapter 5

Conclusions and further work.

> *"The unknown is rushing in on top of us every moment. As I see it, my job is to keep myself open to these collisions, to watch out for all these mysterious goings-on in the world ".*
>
> Paul Auster, *The Art of Hunger*

5.1 Conclusions.

In this work we developed the alternative approach of the censored regression model, with emphasis on Tobit type models. We have shown how the analytical calculations combined with simulations can be used to organized a systematic study of econometric applications of these models.

The re-parametrization used shows that in our case, the global maxima is also achieved. We have illustrated the ideas in the context of models with censoring on the left of zero, and real valued responses, which can also be extended to many other econometric models.

Based on a robustness arguments, Student t models make possible the analysis and estimation with no significant difficulty.

The performance of Newton type algorithms allowed us to get adequate approximations of the regression parameters.

The calculations based on our simulated data in Chapter 4 have been enormously useful for supervising how the degrees of freedom of the model can be estimated.

The information matrix was analytically obtained, throw straightforward derivatives calculations, under the assumption of unknown ν parameter.

Special features that includes different ways of estimate the t-censored model, and how to conduct the statistical inference throw non normal assumptions, and rather general conditions, are considered as future work. Part of these issues are briefly discussed in the following sections.

5.2 The scale mixture representation of the Student-t distribution.

As it is well known, the t-distribution can be represented as scale mixture of the normal distribution. By exploring this fact, a EM algorithm type could be explored to obtain the maximum likelihood estimators. That is, considering that an equivalent specification of the model, $y_i^* \stackrel{ind.}{\sim} t(x_i^T\beta, \sigma^2, \nu)$, $i = 1, \ldots, N$, is given by

$$y_i^*|v_i \stackrel{ind.}{\sim} \mathcal{N}(x_i^T\beta, v_i^{-1}\sigma^2) \quad \text{and} \quad v_i \stackrel{iid}{\sim} Gamma(\nu/2, \nu/2),$$

$i = 1, \ldots, N$. Hence, by augmenting the data ($y_i = \max\{0, y_i^*\}; i = 1, \ldots, N$) with the missing quantities ($v_i; i = 1, \ldots, N$), we obtain the following *completed* log-likelihood function:

$$\begin{aligned} \log L_C(\theta) &= \sum_{i \in I_0} \log\left(1 - \Phi\left(\frac{\sqrt{v_i}\, x_i^T\beta}{\sigma}\right)\right) + N_1 \log\sigma - \frac{1}{2}\sum_{i \in I_1} v_i z_i^2 \\ &+ \sum_{i=1}^{N} \left\{\frac{1}{2}\log v_i + \log g(v_i)\right\}, \end{aligned} \tag{5.1}$$

where as before $z_i = (y_i - x_i^T\beta)/\sigma$, and $g(v)$ is the $Gamma(\nu/2, \nu/2)$-pdf, i.e.:

$$\log g(v) = \frac{\nu}{2}\log\left(\frac{\nu}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \left(\frac{\nu}{2} - 1\right)\log v - \frac{\nu}{2}v. \tag{5.2}$$

Considering (5.1) the following EM scheme can be implemented to compute the MLE:

(i) Calculate the function: $Q(\theta, \theta_k) = E[\log L_C(\theta_k) \mid y_1, \ldots, y_N]$

(ii) Maximize $Q(\theta, \theta_k)$ with respect to θ.

In the context of our Tobit model, we should define the information that we might consider as latent, in order to simplify the resulting expressions. The variables may be written in the following representation:

$$Y_i = Z_i \mathbf{1}\{Z_i \geq 0\}, \qquad Z_i \mid x_i \sim t_\nu(x_i^T\beta, \sigma^2).$$

So, what we actually need to apply this EM scheme is:

(i) Find the likelihood of (x_i, Y_i) and then (x_i, Y_i, Z_i).

(ii) Calculate the function: $Q(\theta, \theta^k) = E[\log L^C(\theta^k) \mid x_1, Y_1, \ldots, x_n, Y_n]$

5.3 Further extensions of the censored regression model.

Further work is steel needed in order to apply these models to new and existing problems, and also into a theoretical development work.

In first place, let's consider *presence of measurement error.* Typically, it is assumed in Tobit models that the exogenous variables are bounded constants and are exactly observed. However, in many real problems, such assumptions are not always appropriate, and in most cases may result in inaccurate and inconsistent estimates. In this context, several authors have considered the censored regression model with presence of measurement errors. Wang (1998) considers the censored regression model with errors-in-variables as

$$
\begin{aligned}
y_i^* &= x_i^T \beta + u_i \\
y_i &= \max\{y_i^*, 0\} \\
x_i &= \xi_i + v_i, \qquad (5.3)
\end{aligned}
$$

where $\xi_i \in \mathbb{R}^k$ are the unobserved variables; β_0, β_1 the regression coefficients, y_i, x_i the observed variables, and u_i, v_i the errors. In his work, Wang considered that u_i, v_i, and ξ_i are independently and normally distributed with means 0, 0, μ_ξ and variances $\sigma_u, \Sigma_v, \Sigma_\xi$, respectively.

As we know, the presence of measurement error causes many difficulties and complexities in the statistical inference process of the model, basically because now the x_i are no longer constants and therefor its distribution must be included in the likelihood function of the model.

The results of Olsen (1978) concerning the existence of the unique global maximum likelihood estimate (MLE) for the model (2.1), and the results of Amemiya (1973) concerning the asymptotic normality of the MLE can be used to obtain the corresponding results for the model in (5.3). The log-likelihood function of model (5.3) is, up to a constant

$$\log L = \sum_0 \Phi\left(-\frac{\gamma_0 + x_i^T\gamma_1}{\sqrt{\sigma_w}}\right) - \frac{N_1}{2}\log\sigma_w - \frac{1}{2\sigma_w}\sum_1 (y_i - \gamma_0 - x_i^T\gamma_1)^2$$
$$- \frac{N}{2}\log\det\Sigma_x - \frac{1}{2}\sum_{i=1}^{N}(x_i - \mu_x)^T\Sigma_x^{-1}(x_i - \mu_x). \qquad (5.4)$$

where $(\gamma_0, \gamma_1, \sigma_w, \mu_x, \Sigma_x)$ are related to the original set of parameters $(\beta_0, \beta_1, \sigma_u, \mu_\xi, \Sigma_\xi, \Sigma_v)$ throw the identities

$$\beta_0 = \gamma_0 - \mu_x^T\Delta\gamma_1 \quad , \quad \beta_1 = (I + \Delta)\gamma_1, \sigma_u = \sigma_w - \gamma_1^T\Sigma_x\Delta\gamma_1,$$
$$\mu_\xi = \mu_x \quad , \quad \Sigma_\xi = \Sigma_x(I + \Delta)^{-1},$$

where $\Delta = \Sigma_\xi^{-1}\Sigma_v$.

Because of the uniqueness mentioned above, the maximization of (5.4) can be obtained using standard numerical methods such as Newton-Raphson. First and second derivatives needed for this procedure are given in Wang, 1998. Potential work is based on the same model proposed by Wang, with the assumption of Student-t distribution of the disturbances u_i, $i = 1, \ldots, N$, and keeping v_i, x_i independent and normally distributed, $i = 1, \ldots, N$.

Bayesian framework.

Many applications of censored regression models are not just of frequentist's interest; there's also simulation based methods (among some others) developed by Bayesians (Chib & Greenberg 1995, Geweke 1989, Chib 1992), which are strongly related to Markov chain Monte Carlo (MCMC) techniques. In the particular case of Tobit models, the Bayesian approach focus on the posterior inference of the regression coefficients.

Let's consider the Tobit model of Tobin (1958), in which the observation is generated by the model in (2.1). Once the likelihood function of this model in (2.6) is multiplied by the prior density, is difficult to simplify for use in the Gibbs sampling algorithm (Geman & Geman, 1984; Gelfand & Smith, 1990). In this context, the parameter space is enlarged considering as latent data those observations which are censored. To see this, notice that $z = (z_i), i \in I_0$, with I_0 as defined in Section 3.2, is the vector containing the latent data, and $Y_N = (y_1, \ldots, y_N)$ is the observed data.

Thus, the Gibbs sampler apply to the blocks β, σ^2, and z, is based on the calculations of the full conditional distribution of $\beta \mid Y_N, z, \sigma^2$, $\sigma^2 \mid Y_N, z, \beta$, and $z \mid Y_N, \beta, \sigma^2$. This distribution are all tractable and the Gibbs sampling is readily applied. (For more details about these distributions, see Chib & Greenberg 1995). This MCMC approach can be straightly modified with the independent Student-t link with ν degrees of freedom (see Albert & Chib 1993a).

Finally, another important topic is to consider the censored regression model with *skew-elliptical* distribution of the error terms. This kind of approach generally provides more flexibility and more robust estimates of the parameters in the model. Distributions such the Skew-t, and the Skew-normal are typically used to fit multiple regression models. In order to extend, and also to compare performance and behavior of censored regression models, further work is required, proposing different types of error terms distribution. The work will focus again on the estimation and the information matrix, which are basic inferential matters needed to overcome a complete study.

Appendix A

Some results on truncated distributions.

A.1 Moments of the truncated normal distribution.

Suppose the random variable X is $N(0,1)$, and we consider the distribution of X given $X \geq c_1$. The mean and variance of this truncated distribution are given by

$$\begin{aligned} E(X \mid X \geq c_1) &= \frac{\phi(c_1)}{1-\Phi(c_1)} = M_1 \\ Var(X \mid X \geq c_1) &= 1 - M_1(M_1 - c_1). \end{aligned}$$

If the truncation is from above, so that we consider the distribution of X given $X \leq c_2$, then

$$\begin{aligned} E(X \mid X \leq c_2) &= \frac{-\phi(c_2)}{\Phi(c_2)} = M_2 \\ Var(X \mid X \leq c_2) &= 1 - M_2(M_2 - c_2). \end{aligned}$$

If the distribution is double truncated, so that we consider $c_1 \leq X \leq c_2$, then

$$\begin{aligned}E(X \mid c_1 \leq X \leq c_2) &= \frac{\phi(c_1) - \phi(c_2)}{\Phi(c_2) - \Phi(c_1)} = M \\ Var(X \mid c_1 \leq X \leq c_2) &= 1 - M^2 + \frac{c_1\phi(c_1) - c_2\phi(c_2)}{\Phi(c_2) - \Phi(c_1)}.\end{aligned}$$

If X has the normal distribution with mean μ and variance σ^2 (instead of mean 0 and variance 1), then in the preceding formulas we have to replace X, c_1, and c_2 by $(X-\mu)/\sigma$, $(c_1-\mu)/\sigma$, and $(c_2-\mu)/\sigma$, respectively (Johnson and Kotz, 1970, pp. 81-3).

A more general result is given in the following lemma.
Lemma A.1: Let $z \sim t(0, 1, \nu)$. Then for any integrable function g:

$$E\{g(z)|z < c\} = E\{g(-z)|z + c > 0\}.$$

A.2 Some derivatives of $\phi(\cdot)$ and $\Phi(\cdot)$.

Very often we need the derivatives of functions of the form $\Phi(\alpha/\lambda)$ and $\phi(\alpha/\lambda)$ with respect to the parameters α and λ. Using that

$$\begin{aligned}\phi(y) &= (2\pi)^{-1/2}e^{-y^2/2} \\ \Phi(y) &= \int_{-\infty}^{y} \phi(u)du,\end{aligned}$$

it can be verified (using the standard formulas for finding a derivative of an expression with an integral sign) that

$$\begin{aligned}
\frac{\partial \Phi}{\partial \alpha} &= \frac{1}{\lambda}\phi\left(\frac{\alpha}{\lambda}\right) \\
\frac{\partial \Phi}{\partial \lambda} &= -\frac{\alpha}{\lambda^2}\phi\left(\frac{\alpha}{\lambda}\right) \\
\frac{\partial \phi}{\partial \alpha} &= -\frac{\alpha}{\lambda^2}\phi\left(\frac{\alpha}{\lambda}\right) \\
\frac{\partial \phi}{\partial \lambda} &= \frac{\alpha^2}{\lambda^3}\phi\left(\frac{\alpha}{\lambda}\right)
\end{aligned}$$

A.3 The Student-t Distribution.

The Student distribution with location parameter μ, scale parameter σ^2, and ν degrees of freedom can be defined as the distribution of the random variable $Z = X/\sqrt{Y/\nu}$, where X and Y are independent random variables with $X \sim N(\mu, \sigma^2)$ and $Y \sim \chi^2_\nu$. The notation $Z \sim t(\mu, \sigma^2, \nu)$ is used to refer this distribution.

It is say that a random variable Z has standard Student-t distribution with ν degrees of freedom, now denoted by $t(0, 1, \nu)$, if its density function is

$$t(z; \nu) = c(\nu)\left(1 + \frac{z^2}{\nu}\right)^{-(\nu+1)/2}, \quad -\infty < z < \infty,$$

where

$$c(\nu) \quad = \quad \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{1}{2}\right)} \cdot \nu^{-1/2}.$$

Note that $t(-z;\nu) = t(z;\nu)$, and $t'(z;\nu) = -\dfrac{zt(z;\nu)(\nu+1)}{\nu+z^2}$. The associated cumulative distribution function is:

$$T(Z;\nu) = P(Z \leq z) \quad = \quad \int_{-\infty}^{z} t(w;\nu)dw.$$

Note that $T'(z;\nu) = t(z;\nu)$ and $T(-z;\nu) = 1 - T(z;\nu)$.

Letting $Y = \mu + \sigma Z$, we get:

$$P(Y \leq y) = P(\mu + \sigma Z \leq y) \quad = \quad P\left(Z \leq \frac{y-\mu}{\sigma}\right) = T\left(\frac{y-\mu}{\sigma};\nu\right).$$

Hence, the density of Y is

$$f_Y(y) = \frac{1}{\sigma}t\left(\frac{y-\mu}{\sigma};\nu\right) \quad = \quad \frac{1}{\sigma}c(\nu)\left(1+\frac{1}{\nu}\left(\frac{y-\mu}{\sigma}\right)^2\right)^{-(\nu+1)/2}, \quad -\infty < y < \infty.$$

This distribution is called t-distribution with location parameter μ, scale parameter σ^2, and ν degrees of freedom. The notation $Y \sim t(\mu, \sigma^2, \nu)$ is used in such case.

Another stochastic representation of $Y \sim t(\mu, \sigma^2, \nu)$, is given by

$$Y \stackrel{\mathcal{L}}{=} \mu + \sigma V^{-1/2} X,$$

where $V \sim Gamma(\nu/2, \nu/2)$, and is independent of $X \sim N(0,1)$. From this stochastic representation, is easy to show that

$$\begin{aligned} E(Y) &= \mu, \quad \nu > 1 \\ Var(Y) &= \frac{\nu}{\nu - 2}\sigma^2, \quad \nu > 2. \end{aligned}$$

A.4 Truncated Student-t Distribution

Let $Y \sim t(\mu, \sigma^2, \nu)$. The truncated distribution of Y on the set $A = [a_1, a_2]$, $-\infty < a_1 < a_2 < \infty$, is given by

$$f_{Y|Y\in A}(y) = \begin{cases} \frac{1}{\sigma}\frac{t(z;\nu)}{T(\alpha_2;\nu) - T(\alpha_2;\nu)}, & \text{for } y \in A \\ 0, & \text{in other case,} \end{cases}$$

where $z = (y - \mu)/\sigma$, and $\alpha_i = (a_i - \mu)/\sigma$, $i = 1, 2$.

The mean of Y given $Y \in A$ is

$$E(Y \mid Y \in A) = \mu + \frac{\sigma}{\nu - 1}\frac{t(\alpha_2;\nu)(\nu + \alpha_2^2) - t(\alpha_1;\nu)(\nu + \alpha_1^2)}{T(\alpha_2;\nu) - T(\alpha_1;\nu)}, \quad \nu > 1.$$

Letting $a_2 \to \infty$, we have

$$E(Y \mid Y \geq a_1) = \mu + \sigma \left(\frac{\nu + \alpha_1^2}{\nu - 1} \right) R(\alpha_1; \nu),$$

where $R(\alpha; \nu) = t(\alpha; \nu)/T(-\alpha; \nu)$ is known as the failure or hazard function, and in the normal case with $\nu \to \infty$ is called inverse Mill's ratio.

Noticing that the rate $\left(\frac{\nu+\alpha_1^2}{\nu-1}\right) > 0$, provided $\nu > 1$, we have that $E(Y) < E(Y \mid Y \in [a_1, \infty))$. Analogously,

$$E(Y \mid Y \in (-\infty, a_2]) \;=\; \mu - \sigma \left(\frac{\nu + \alpha_2^2}{\nu - 1} \right) R(\alpha_2; \nu)$$

For $Z \sim t(0, 1, \nu)$, a more general result can be obtained by using the scale mixture representation

$$Z \stackrel{\mathcal{L}}{=} V^{-1/2} X,$$

where $V \sim Gamma(\nu/2, \nu/2)$, and $X \sim N(0, 1)$. Thus, by letting $Z_{(a,b)} \stackrel{\mathcal{L}}{=} Z \mid a < Z < b$, and $m_k(Z_{(a,b)}) = E(Z^k \mid a < Z < b)$, we have:

$$m_k(Z_{(a,b)}) \;=\; \frac{E_V \left\{ V^{-k/2} \left(T(\sqrt{V} b; \nu) - T(\sqrt{V} a; \nu) \right) E_{X|V} \left(X^k \mid \sqrt{V} a < X < \sqrt{V} b \right) \right\}}{T(b; \nu) - T(a; \nu)},$$

provided $\nu > k$.

Given $a = 0$ and $b = \infty$, the above expression for the truncated moments become:

$$m_k(Z_{(0,\infty)}) = E_V\left\{V^{-k/2}E_{X|V}\left(X^k \mid 0 < X < \infty\right)\right\}$$
$$= E(V^{-k/2})E\left(X^k \mid 0 < X < \infty\right),$$

with:

$$E(V^{-k/2}) = \frac{\Gamma\left(\frac{\nu-k}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}$$
$$E\left(X^k \mid 0 < X < \infty\right) = \frac{\partial^k}{\partial t^k}M_{X_{(0,\infty)}}(t)|_{t=0},$$

where $M_{X_{(a,b)}}$ is given in the following lemma:

Lemma 1: *Let $X \sim \mathcal{N}(0,1)$. The moment generating function for the random variable $X_{(a,b)} \overset{\mathcal{L}}{=} X \mid a < X < b$* is:

$$M_{X_{(a,b)}}(t) = M_X(t)\frac{\Phi(b-t)-\Phi(a-t)}{\Phi(b)-\Phi(a)},$$

where $M_X(t) = e^{t^2/2}$.

A.5 Further properties of the t-distributions and its truncations parents.

Lemma A.2: Let $t(z;\nu)$ be the $t(0,1,\nu)$-pdf. Then:

$$\left(1+\frac{z^2}{\nu}\right)^{-m/2} t(z;\nu) = \frac{t(0;\nu)}{t(0;\nu+m)}t\left(\sqrt{\frac{\nu+m}{\nu}}\,z;\nu+m\right).$$

Lemma A.3: Let $z_m \sim t(0, 1, \nu + m)$*, with* $z = z_0$*. Then:*

$$E\left\{\left(1+\frac{z^2}{\nu}\right)^{-m/2} z^k | z + c > 0\right\} = \frac{t(0;\nu)}{t(0;\nu+m)}\left(\frac{\nu}{\nu+m}\right)^{(k+1)/2}\frac{T(c_m;\nu+m)}{T(c;\nu)}m_k(c_m;\nu+m),$$

where $m_k(c_m; \nu + m) = E\{z_m^k | z_m + c_m > 0\}$*,* $c_m = \sqrt{\frac{\nu+m}{\nu}}\,c$ *and* $T(\cdot; \nu + m)$ *denotes the cdf of* z_m*.*

Lemma A.4: Let $m_k(c; \nu) = E\{z^k | z + c > 0\}$*,* $\nu > k$*, where* $z \sim t(0, 1, \nu)$*. Then, for* $k = 1, 2, 3, 4$ *it follows that:*

$$\begin{aligned}
m_1(c;\nu) &= \frac{\nu}{\nu-1}\left(1+\frac{c^2}{\nu}\right)r(c;\nu), \quad \nu > 1,\\
m_2(c;\nu) &= \frac{\nu}{\nu-2}\frac{T\left(c_{-2};\nu-2\right)}{T(c;\nu)} - c\,m_1(c;\nu), \quad \nu > 2,\\
m_3(c;\nu) &= \frac{2\nu^2}{(\nu-1)(\nu-3)}\left(1+\frac{c^2}{\nu}\right)^2 r(c;\nu) + c^2 m_1(c;\nu), \quad \nu > 3,\\
m_4(c;\nu) &= \frac{3\nu^2}{(\nu-2)(\nu-4)}\frac{T\left(c_{-4};\nu-4\right)}{T(c;\nu)} - 3c\,m_3(c;\nu) + 2c^3 m_1(c;\nu), \quad \nu > 4,
\end{aligned}$$

where $r(c;\nu) = \frac{t(c;\nu)}{T(c;\nu)}$ *and* $c_{-k} = \sqrt{\frac{\nu-k}{\nu}}\,c$*,* $\nu > k$*.*

The proofs of Lemma 1 to Lemma 3 are straightforward, and for the proof of Lemma 4 see Arellano-Valle and Genton (2008). Finally, from the Lemmas 2 and 4 we obtain following necessary truncated moments for the computation of the expected information matrix of the t censored model:

Lemma A.5: Let $z_m \sim t(0,1,\nu+m)$, with $z_0 = z$. Then:

$$
\begin{aligned}
E\left\{\left(1+\frac{z^2}{\nu}\right)^{-1}|z+c>0\right\} &= \left(\frac{\nu}{\nu+1}\right)\frac{T(c_2;\nu+2)}{T(c;\nu)} \\
E\left\{\left(1+\frac{z^2}{\nu}\right)^{-1}z|z+c>0\right\} &= \left(\frac{\nu}{\nu+1}\right)\left(\frac{\nu}{\nu+2}\right)^{1/2}\frac{T(c_2;\nu+2)}{T(c;\nu)}m_1(c_2;\nu+2) \\
&= \left(\frac{\nu}{\nu+1}\right)r(c;\nu) \\
E\left\{\left(1+\frac{z^2}{\nu}\right)^{-1}z^2|z+c>0\right\} &= \left(\frac{\nu}{\nu+1}\right)\left(\frac{\nu}{\nu+2}\right)\frac{T(c_2;\nu+2)}{T(c;\nu)}m_2(c_2;\nu+2) \\
&= \left(\frac{\nu}{\nu+1}\right)\{1-c\,r(c;\nu)\} \\
E\left\{\left(1+\frac{z^2}{\nu}\right)^{-2}z^2|z+c>0\right\} &= \left(\frac{\nu}{\nu+1}\right)\left(\frac{\nu+2}{\nu+3}\right)\left(\frac{\nu}{\nu+4}\right)\frac{T(c_4;\nu+4)}{T(c;\nu)}m_2(c_4;\nu+4) \\
&= \left(\frac{\nu}{\nu+1}\right)\left(\frac{\nu}{\nu+3}\right)\frac{T(c_2;\nu+2)}{T(c;\nu)}\{1-c_2\,r(c_2;\nu+2)\} \\
E\left\{\left(1+\frac{z^2}{\nu}\right)^{-2}z^3|z+c>0\right\} &= \left(\frac{\nu}{\nu+1}\right)\left(\frac{\nu+2}{\nu+3}\right)\left(\frac{\nu}{\nu+4}\right)^{3/2}\frac{T(c_4;\nu+4)}{T(c;\nu)}m_3(c_4;\nu+4) \\
&= \left(\frac{\nu}{\nu+1}\right)\left(\frac{\nu}{\nu+3}\right)\Bigg\{2\,r(c;\nu) \\
&+ \left(\frac{\nu}{\nu+2}\right)^{1/2}\frac{T(c_2;\nu+2)}{T(c;\nu)}c_2^2\,r(c_2;\nu+2)\Bigg\} \\
E\left\{\left(1+\frac{z^2}{\nu}\right)^{-2}z^4|z+c>0\right\} &= \left(\frac{\nu}{\nu+1}\right)\left(\frac{\nu+2}{\nu+3}\right)\left(\frac{\nu}{\nu+4}\right)^{2}\frac{T(c_4;\nu+4)}{T(c;\nu)}m_4(c_4;\nu+4) \\
&= \left(\frac{\nu}{\nu+1}\right)\left(\frac{\nu}{\nu+3}\right)\Bigg\{3-6c\,r(c;\nu) \\
&- \left(\frac{\nu}{\nu+2}\right)\frac{T(c_2;\nu+2)}{T(c;\nu)}c_2^3\,r(c_2;\nu+2)\Bigg\}
\end{aligned}
$$

where as before $c_k = \sqrt{\frac{\nu+k}{\nu}}\,c$.

Appendix B

Derivatives.

B.1 The first derivatives $T(c;\nu)$ with respect to ν

Using that $\partial T(c;\nu)/\partial\nu = \int_{-\infty}^{c} t(z;\nu)[\partial \log t(z;\nu)/\partial\nu]dz$, we have

$$\begin{aligned}
\frac{\partial T\left(c;\nu\right)}{\partial\nu} &= \frac{1}{2}\left\{\psi\left(\frac{\nu+1}{2}\right)-\psi\left(\frac{\nu}{2}\right)-\frac{1}{\nu}\right\}T\left(c;\nu\right)-\frac{1}{2}\int_{-\infty}^{c}t(z;\nu)\log\left(1+\frac{z^2}{\nu}\right)dz \\
&+ \frac{1}{2}\left(\frac{\nu+1}{\nu}\right)\int_{-\infty}^{c}\frac{z^2}{\nu}\left(1+\frac{z^2}{\nu}\right)^{-1}t(z;\nu)dz \\
&= \frac{1}{2}\left\{\psi\left(\frac{\nu+1}{2}\right)-\psi\left(\frac{\nu}{2}\right)-\frac{1}{\nu}\right\}T\left(c;\nu\right)-\frac{1}{2}\int_{-\infty}^{c}t(z;\nu)\log\left(1+\frac{z^2}{\nu}\right)dz \\
&+ \frac{1}{2}\frac{1}{\nu+2}\int_{-\infty}^{\sqrt{\frac{\nu+2}{\nu}}c}z^2t\left(z;\nu+2\right)dz \\
&= \frac{1}{2}\left\{\psi\left(\frac{\nu+1}{2}\right)-\psi\left(\frac{\nu}{2}\right)-\frac{1}{\nu}\right\}T\left(c;\nu\right)-\frac{1}{2}\int_{-\infty}^{c}t(z;\nu)\log\left(1+\frac{z^2}{\nu}\right)dz \\
&+ \frac{1}{2}\frac{1}{\nu}\{1-cr(c;\nu)\}T\left(c;\nu\right),
\end{aligned}$$

since by Lemma A.1

$$\left(1+\frac{z^2}{\nu}\right)^{-1}t(z;\nu) = \left(\frac{\nu}{\nu+1}\right)\sqrt{\frac{\nu+2}{\nu}}\,t\left(\sqrt{\frac{\nu+2}{\nu}}\,z;\nu+2\right)$$

and by Lemma A.4

$$\int_{-\infty}^{\sqrt{\frac{\nu+2}{\nu}}\,c} z^2 t\,(z;\nu+2)\,dz = \left(\frac{\nu+2}{\nu}\right) T\,(c;\nu)\,\{1-cr(c;\nu)\}.$$

B.2 Second derivatives

Considering again that $\partial z_i/\partial\gamma^T = (\partial z_i/\partial\gamma)^T = -x_i^T$ and $\partial z_i/\partial\tau = y_i = (z_i + x_i^T\gamma)/\tau$, we obtain from (3.8)-(3.10) that

$$\begin{aligned}
S_{\gamma\gamma} &= \sum_{i\in I_0}[R'(x_i^T\gamma;\nu)r(x_i^T\gamma;\nu) + R(x_i^T\gamma;\nu)r'(x_i^T\gamma;\nu)]x_i x_i^T \\
&- \left(\frac{\nu+1}{\nu}\right)\sum_{i\in I_1}\left\{\left(1+\frac{z_i^2}{\nu}\right)^{-1} - 2\frac{z_i^2}{\nu}\left(1+\frac{z_i^2}{\nu}\right)^{-2}\right\} x_i x_i^T \\
S_{\gamma\tau} &= \left(\frac{\nu+1}{\nu}\right)\sum_{i\in I_1}\left\{\left(1+\frac{z_i^2}{\nu}\right)^{-1} - 2\frac{z_i^2}{\nu}\left(1+\frac{z_i^2}{\nu}\right)^{-2}\right\} y_i x_i \\
S_{\gamma\nu} &= -\sum_{i\in I_0}\left\{\frac{\partial R(x_i^T\gamma;\nu)}{\partial\nu} r(x_i^T\gamma;\nu) + R(x_i^T\gamma;\nu)\frac{\partial r(x_i^T\gamma;\nu)}{\partial\nu}\right\} x_i \\
&- \frac{1}{\nu}\sum_{i\in I_1}\left\{\frac{z_i}{\nu}\left(1+\frac{z_i^2}{\nu}\right)^{-1} - \left(\frac{\nu+1}{\nu}\right)\frac{z_i^3}{\nu}\left(1+\frac{z_i^2}{\nu}\right)^{-2}\right\} x_i \\
S_{\tau\tau} &= -\frac{N_1}{\tau^2} - \left(\frac{\nu+1}{\nu}\right)\sum_{i\in I_1}\left\{\left(1+\frac{z_i^2}{\nu}\right)^{-1} - 2\frac{z_i^2}{\nu}\left(1+\frac{z_i^2}{\nu}\right)^{-2}\right\} y_i^2 \\
S_{\tau\nu} &= \frac{1}{\nu}\sum_{i\in I_1}\left\{\frac{z_i}{\nu}\left(1+\frac{z_i^2}{\nu}\right)^{-1} - \left(\frac{\nu+1}{\nu}\right)\frac{z_i^3}{\nu}\left(1+\frac{z_i^2}{\nu}\right)^{-2}\right\} y_i,
\end{aligned}$$

where $R'(z;\nu) = \partial R(z;\nu)/\partial z = (1+R(z;\nu))R(z;\nu)r(z;\nu)$ and $r'(z;\nu) = \partial r(z;\nu)/\partial z = [t'(z;\nu)/T(z;\nu)] - [r(z;\nu)]^2$, with $t'(z;\nu) = \partial t(z;\nu)/\partial z = t(z;\nu)[\partial \log t(z;\nu)/\partial z]$ given by (see also Lemma A.1)

$$t'(z;\nu) = -\left(\frac{\nu+1}{\nu}\right) z\left(1+\frac{z^2}{\nu}\right)^{-1} t(z;\nu) = -\sqrt{\frac{\nu+2}{\nu}}\, zt\left(\sqrt{\frac{\nu+2}{\nu}}\, z;\nu+2\right);$$

and

$$
\begin{aligned}
S_{\nu\nu} &= \frac{1}{4}N_1\left\{\psi'\left(\frac{\nu+1}{2}\right)-\psi'\left(\frac{\nu}{2}\right)+\frac{2}{\nu^2}\right\}\\
&- \frac{1}{2}\sum_{i\in I_0}\frac{\partial R(x_i^T\gamma;\nu)}{\partial\nu}\left\{\psi\left(\frac{\nu+1}{2}\right)-\psi\left(\frac{\nu}{2}\right)-b_{01}(x_i^T\gamma;\nu)-\frac{1}{\nu}(x_i^T\gamma)r_i(x_i^T\gamma;\nu)\right\}\\
&- \frac{1}{2}\sum_{i\in I_0}R(x_i^T\gamma;\nu)\left\{\frac{1}{2}\psi'\left(\frac{\nu+1}{2}\right)-\frac{1}{2}\psi'\left(\frac{\nu}{2}\right)-\frac{\partial b_{01}(x_i^T\gamma;\nu)}{\partial\nu}+\frac{1}{\nu^2}(x_i^T\gamma)r_i(x_i^T\gamma;\nu)\right.\\
&- \left.\frac{1}{\nu}(x_i^T\gamma)\frac{\partial r_i(x_i^T\gamma;\nu)}{\partial\nu}\right\}+\frac{1}{2}\frac{1}{\nu}\sum_{i\in I_1}\frac{z_i^2}{\nu}\left(1+\frac{z_i^2}{\nu}\right)^{-1}-\frac{1}{2}\frac{1}{\nu^2}\sum_{i\in I_1}\frac{z_i^2}{\nu}\left(1+\frac{z_i^2}{\nu}\right)^{-1}\\
&+ \frac{1}{2}\frac{1}{\nu}\left(\frac{\nu+1}{\nu}\right)\sum_{i\in I_1}\left\{-\frac{z_i^2}{\nu}\left(1+\frac{z_i^2}{\nu}\right)^{-1}+\left(\frac{z_i^2}{\nu}\right)^2\left(1+\frac{z_i^2}{\nu}\right)^{-2}\right\},
\end{aligned}
$$

where, by applying appropriately the results in Lemmas A.1-A.5,

$$
\begin{aligned}
\frac{\partial R(x_i^T\gamma;\nu)}{\partial \nu} &= \{1+R(x_i^T\gamma;\nu)\}^2 \frac{\partial T(x_i^T\gamma;\nu)}{\partial \nu} \\
&= \frac{1}{2}\{1+R(x_i^T\gamma;\nu)\}^2 \left\{\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - b_{01}(x_i^T\gamma;\nu) - \frac{1}{\nu}(x_i^T\gamma) r(x_i^T\gamma;\nu)\right\} \\
\frac{\partial r(x_i^T\gamma;\nu)}{\partial \nu} &= r(x_i^T\gamma;\nu)\left\{\frac{\partial \log t(x_i^T\gamma;\nu)}{\partial \nu} - \frac{1}{T(x_i^T\gamma;\nu)}\frac{\partial T(x_i^T\gamma;\nu)}{\partial \nu}\right\} \\
&= \frac{1}{2} r(x_i^T\gamma;\nu)\left\{-\log\left(1+\frac{(x_i^T\gamma)^2}{\nu}\right) + \left(\frac{\nu+1}{\nu}\right)\frac{(x_i^T\gamma)^2}{\nu}\left(1+\frac{(x_i^T\gamma)^2}{\nu}\right)^{-1}\right. \\
&+ \left. b_{01}\left(x_i^T\gamma;\nu\right) - \frac{1}{\nu}\{1-(x_i^T\gamma) r(x_i^T\gamma;\nu)\}\right\} \\
\frac{\partial b_{01}(x_i^T\gamma;\nu)}{\partial \nu} &= -b_{01}(x_i^T\gamma;\nu)\frac{1}{T(x_i^T\gamma;\nu)}\frac{\partial T(x_i^T\gamma;\nu)}{\partial \nu} \\
&+ \frac{1}{T(x_i^T\gamma;\nu)}\int_{-\infty}^{x_i^T\gamma}\left\{-\frac{z^2}{\nu^2}\left(1+\frac{z^2}{\nu}\right)^{-1} + \log\left(1+\frac{z^2}{\nu}\right)\frac{\partial \log t(z;\nu)}{\partial \nu}\right\} t(z;\nu)dz \\
&= \frac{1}{2}\{b_{01}\left(x_i^T\gamma;\nu\right)\}^2 - \frac{1}{2} b_{02}(x_i^T\gamma;\nu) \\
&- \frac{1}{2}\frac{1}{\nu}\{1-(x_i^T\gamma) r(x_i^T\gamma;\nu)\} b_{01}(x_i^T\gamma;\nu) - \frac{1}{\nu(\nu+1)}\{1-(x_i^T\gamma) r(x_i^T\gamma;\nu)\} \\
&+ \frac{1}{2}\frac{1}{\nu+2}\frac{T\left(\sqrt{\frac{\nu+2}{\nu}}\, x_i^T\gamma;\nu+2\right)}{T\left(x_i^T\gamma;\nu\right)} b_{21}\left(\sqrt{\frac{\nu+2}{\nu}}\, x_i^T\gamma;\nu+2\right).
\end{aligned}
$$

With these results we obtain the second derivatives given in Section 3. Finally, to compute the necessary ingredients of the expected information matrix, we note from

Lemma A.5 that

$$
\begin{aligned}
E\left\{\left(1+\frac{z_i^2}{\nu}\right)^{-1}-\frac{2}{\nu}z_i^2\left(1+\frac{z_i^2}{\nu}\right)^{-2}\right\} &= \left(\frac{\nu}{\nu+3}\right)\frac{T(c_2;\nu+2)}{T(c;\nu)} \\
&\times \left\{1+\frac{2}{\nu+1}c_2\, r(c_2;\nu+2)\right\} \\
E\left\{z_i\left(1+\frac{z_i^2}{\nu}\right)^{-1}-\frac{2}{\nu}z_i^3\left(1+\frac{z_i^2}{\nu}\right)^{-2}\right\} &= \frac{\nu(\nu-1)}{(\nu+1)(\nu+3)}r(c;\nu)-\frac{2\nu}{(\nu+1)(\nu+3)} \\
&\times \left(\frac{\nu}{\nu+2}\right)^{1/2}\frac{T(c_2;\nu+2)}{T(c;\nu)}c_2^2\, r(c_2;\nu+2) \\
E\left\{z_i\left(1+\frac{z_i^2}{\nu}\right)^{-1}-\left(\frac{\nu+1}{\nu}\right)z_i^3\left(1+\frac{z_i^2}{\nu}\right)^{-2}\right\} &= -\frac{\nu(\nu-1)}{(\nu+1)(\nu+3)}r(c;\nu)-\left(\frac{\nu}{\nu+3}\right) \\
&\times \left(\frac{\nu}{\nu+2}\right)^{1/2}\frac{T(c_2;\nu+2)}{T(c;\nu)}c_2^2\, r(c_2;\nu+2) \\
E\left\{z_i^2\left(1+\frac{z_i^2}{\nu}\right)^{-1}-\frac{2}{\nu}z_i^4\left(1+\frac{z_i^2}{\nu}\right)^{-2}\right\} &= \frac{\nu(\nu-3)}{(\nu+1)(\nu+3)}-\frac{\nu(\nu-9)}{(\nu+1)(\nu+3)}c\, r(c;\nu) \\
&+ \frac{2\nu^2}{(\nu+1)(\nu+2)(\nu+3)}\frac{T(c_2;\nu+2)}{T(c;\nu)} \\
&\times c_2^3\, r(c_2;\nu+2) \\
E\left\{z_i^2\left(1+\frac{z_i^2}{\nu}\right)^{-1}-\left(\frac{\nu+1}{\nu}\right)z_i^4\left(1+\frac{z_i^2}{\nu}\right)^{-2}\right\} &= -\frac{2\nu^2}{(\nu+1)(\nu+3)}+\frac{\nu(5\nu+3)}{(\nu+1)(\nu+3)}c\, r(c;\nu) \\
&+ \frac{\nu^2}{(\nu+2)(\nu+3)}\frac{T(c_2;\nu+2)}{T(c;\nu)}c_2^3\, r(c_2;\nu+2) \\
E\left\{2z_i^2\left(1+\frac{z_i^2}{\nu}\right)^{-1}-\left(\frac{\nu+1}{\nu}\right)z_i^4\left(1+\frac{z_i^2}{\nu}\right)^{-2}\right\} &= -\frac{\nu(\nu-3)}{(\nu+1)(\nu+3)}+\frac{\nu^2}{(\nu+1)(\nu+3)}c\, r(c;\nu) \\
&+ \frac{\nu^2}{(\nu+2)(\nu+3)}\frac{T(c_2;\nu+2)}{T(c;\nu)}c_2^3\, r(c_2;\nu+2).
\end{aligned}
$$

Under I_1, we need truncated expectation, the form of $E\left\{z^q\left(1+\frac{z^2}{\nu}\right)^{-s}|z<c\right\}$, where $z\sim t(0,1,\nu)$, which can be obtained from the results in section 2. We need also truncated expectations the form of $E\left\{z^q\left(1+\frac{z^2}{\nu}\right)^{-r}\left[\log\left(1+\frac{z^2}{\nu}\right)\right]^s|z<c\right\}$, which

need to be computed numerically.

Appendix C

R routines and functions.

C.0.1 Fitting the normal model to Tobin's data set.

```
library(survival)
tofit<- survreg(Surv(durable,durable>0,type='left')~age+quant,data=tobin, dist='gaussian')
tobi<-cbind(y<-tobin$durable,x0<-1,x1<-tobin$age,x2<-tobin$quant)
toby <- transform(data.frame(y,x0,x1,x2))
toby[order(y),]
retobi<-toby[order(y),]
cen<-retobi[1:13,]
nocen<-retobi[14:20,]
X1<-cbind(nocen$x0,nocen$x1,nocen$x2)
X0<-cbind(cen$x0,cen$x1,cen$x2)
Y0<-matrix(cen$y)
Y1<-matrix(nocen$y)
t0<-c(15.15,-0.129,-0.045,0.15)
NRcensored<-function(t0){
N1<-13
nu<-7
cont<-1
repeat{
h<-t0[4]
B<-rbind(t0[1],t0[2],t0[3])
f1<-dt(X0%*%B,nu)
f2<-1-pt(X0%*%B,nu)
suma0<-t(X0)%*%(f1/f2)
f1<-(h*Y1-X1%*%B)
f2<-(nu+(h*Y1-X1%*%B)^2)
suma1<-t(X1)%*%(f1/f2)
grad1<--suma0+(nu+1)*suma1
f1<-(h*Y1-X1%*%B)
f2<-(nu+(h*Y1-X1%*%B)^2)
grad2<-N1/h-(nu+1)*t(Y1)%*%(f1/f2)
grad<-c(grad1,grad2)
f1<-(nu-(h*Y1-X1%*%B)^2)
f2<-(nu+(h*Y1-X1%*%B)^2)^2
hh<--N1/(h^2)-(nu+1)*t(Y1^2)%*%(f1/f2)
f1<-(nu-(h*Y1-X1%*%B)^2)
f2<-(nu+(h*Y1-X1%*%B)^2)^2
f3<-matrix(0,nc=1,nr=7)
for(i in 1:7){f3[i,]<-(f1/f2)[i,]*Y1[i,]}
f4<-matrix(0,nc=3,nr=7)
```

```
for(i in 1:7){f4[i,]<-t(X1)[,i]*f3[i,1]}
hBh<-(nu+1)*rbind(sum(f4[,1]),sum(f4[,2]),sum(f4[,3]))
f1<-dt(X0%*%B,nu)
f2<-1-pt(X0%*%B,nu)
f3<-((nu+1)*(X0%*%B)/(nu+(X0%*%B)^2)-(f1/f2))
f4<-0
for(i in 1:13){
f5<-f4+(f1/f2)[i,1]*f3[i,1]*matrix(t(X0)[,1],nr=3,1)%*%matrix(X0[1,],nr=1,3)
f4<-f5
}
suma0<-f4
(nu-(h*Y1-X1%*%B)^2)/(nu+(h*Y1-X1%*%B)^2)
m<-matrix(X1[i,],ncol=1,nrow=3)
suma1<-suma0+(m%*%X1[i,])*esc[1,1]
suma0<-suma2
}
suma2
hBB<-suma1-(nu+1)*suma2
h1<-cbind(hBB,hBh)
h2<-c(hBh,hh)
H<-matrix(rbind(h1,h2),4,4)
theta<-t0-solve(H)%*%grad
print(theta)
dif<-(t(theta-t0))%*%(theta-t0)
if(sqrt(dif)<=0.1) break
t0<-theta
cont<-cont+1
}
return(theta,-solve(h),cont)
}
summary(NRcensored(t0=c(15.15,-0.129,-0.045,0.15)))
```

C.0.2 Fitting the Student-t model, with known ν degrees of freedom to Tobin's data, and simulated data.

```
my.t<-survreg.distributions$t
tfit1<-survreg(Surv(durable,durable>0, type='left')~age + quant,data=tobin,dist=my.t,parms=7)
tfit2<-survreg(Surv(durable,durable>0, type='left')~age + quant,data=tobin,dist=my.t,parms=15)
tfit3<-survreg(Surv(durable,durable>0, type='left')~age + quant,data=tobin,dist=my.t,parms=50)
tfit4<-survreg(Surv(durable,durable>0, type='left')~age + quant,data=tobin,dist=my.t,parms=200)
summary(tfit1); summary(tfit2); summary(tfit3); summary(tfit4)

censoredsample<-function(n){
epsilon<-0.09999
y<-matrix(0,ncol=1,nrow=n)
x<-matrix(0,ncol=1,nrow=n)
for(i in 1:n){
m<-function(x){
1+3*x
}
x[i,1]=runif(1)
#e=rt(n=1,df=12)
e=(1-epsilon)*rnorm(n=1,0,3)+ epsilon*rnorm(n=1,0,3*2)
yast=m(x[i,1])- e
y[i,1]=max(yast,0)
}
list(y=y,x=x)
}
cns<-censoredsample(n=250)
```

```
yt<-cns$y
xt<-cns$x
contnormal<-transform(data.frame(yt,xt))
newtsimulated<-contnormal[order(yt),]
for(i in 1:(length(newtsimulated[,1]))){
if( newtsimulated$yt[i]==0) indice=i
else break;
}
tcen<-newtsimulated[1:indice,]
notcen<-newtsimulated[(indice+1):(length(newtsimulated[,1])),]
X1<-cbind(1,notcen$xt)
X0<-cbind(1,tcen$xt)
Y0<-matrix(tcen$yt)
Y1<-matrix(notcen$yt)
N0<-length(Y0)
N1<-length(Y1)
my.t<-survreg.distributions$t
normfit<-survreg(Surv(yt,yt>0,type='left')~xt,data=cns,dist='gaussian')
tfit1<-survreg(Surv(yt,yt>0,type='left')~xt,data=cns,dist=my.t,parms=7)
tfit2<-survreg(Surv(yt,yt>0,type='left')~xt,data=cns,dist=my.t,parms=15)
tfit3<-survreg(Surv(yt,yt>0,type='left')~xt,data=cns,dist=my.t,parms=50)
tfit4<-survreg(Surv(yt,yt>0,type='left')~xt,data=cns,dist=my.t,parms=200)
summary(normfit)
summary(tfit1)
summary(tfit2)
summary(tfit3)
summary(tfit4)
```

C.0.3 Simulated data, with unknown degrees of freedom.

```
mcestimates<-function(N){
estimates<-matrix(0,ncol=4,nrow=N)
for(k in 1:N){
tcensoredsample<-function(n){
y<-matrix(0,ncol=1,nrow=n)
x<-matrix(0,ncol=1,nrow=n)
for(i in 1:n){
m<-function(x){
1+3*x
}
x[i,1]=runif(1)
e=rt(n=1,df=12)
y[i,1]=max(m(x[i,1])-e,0)
}
list(y=y,x=x)
}
tst<-tcensoredsample(250)
xt<-tst$x
yt<-tst$y
tsimulated<-transform(data.frame(yt,xt))
newtsimulated<-tsimulated[order(yt),]
for(i in 1:(length(newtsimulated[,1]))){
if( newtsimulated$yt[i]==0) indice=i
else break;
}
tcen<-newtsimulated[1:indice,]
notcen<-newtsimulated[(indice+1):(length(newtsimulated[,1])),]
X1<-cbind(1,notcen$xt)
X0<-cbind(1,tcen$xt)
Y0<-matrix(tcen$yt)
```

```
Y1<-matrix(notcen$yt)
N0<-length(Y0)
N1<-length(Y1)
tl<-function(p){
g<-rbind(p[2],p[3])
logv=sum(log(1-pt(X0%*%(g/p[1]),df=p[4])))+
sum(log(dt((Y1-X1%*%g)/p[1],df=p[4])))-N1*log(p[1])
logver<--logv
}
k2<-nlm(tl,c(1,1,3,12))$estimate
estimates[k,]<-k2
}
est<-apply(estimates,2,mean)
list(est=est)
}
results<-mcestimates(1000)
results
nlm(tl,c(1,1,3,2))
nlminb(start=c(1,1,3,2), objective=tl)
```

References

Albert, James H. and Chib, Siddhartha (1993) *Bayesian Analysis of Binary and Polychotomous Response Data.* Journal of the American Statistical Association. Vol. 88, No. 422, Theory and Methods.

Amemiya, Takeshi (1973) *Regression analysis when the dependent variable is truncated normal.* Econometrica, Vol.41, No. 6.

Amemiya, Takeshi (1984) *Tobit models: a survey.* Journal of Econometrics, Vol.24, 2-61.

Amemiya, Takeshi (1985) *Advanced Econometrics.* Harvard University Press. Cambridge, Massachusetts.

Auster, Paul (1988) *The Invention of Solitude.* Publisher: Penguin.

Auster, Paul (1992) *The Art of Hunger.* Sun & Moon Press.

Chib, Siddhartha (1992) *Bayes inference in the Tobit censored regression model.* Journal of Econometrics. 51, 79-99.

Chib, Siddhartha and Greenberg Edward (1995) *Markov Chain Monte Carlo Simulation Methods in Econometrics*. Technical report.

DiCiccio, Thomas J. and Monti, Anna Clara (2008)*Inferential Aspects in the Skew t model*. Technical Report.

Fair, Ray C. (1977) *A note on the computation of the Tobit estimator*. Econometrica, Vol.45, No. 7.

Geweke, John (1993) *Bayesian Treatment of the Independent Student-t Linear Model*. Journal of Applied Econometrics. Vol. 8, S19-S40.

Goldfeld, Stephen M. and Quandt, Richard E. (1976) *Studies in Nonlinear Estimation*. Ballinger Publishing Company. Cambridge, Massachusetts.

Greene, William H. (1981) *On the Asymptotic Bias of the Ordinary Least Squares Estimator of the Tobit Model*. Econometrica, Vol. 49, No. 2.

Lange, Kenneth L., Little, Roderick J.A. and Taylor, Jeremy M.G. (1989)*Robust statistical modeling using the t distribution*. Journal of the American Statistical Association, Vol.84, No.408.

Long, J. Scott (1997) *Regression Models for Categorical and Limited Dependent Variables*. Advanced Quantitative Techniques in the Social Sciences Series. SAGE Publications.

Maddala, G.S. (1983) *Limited-dependent and qualitative variables in econometrics.* Econometric Society Monographs in Quantitative Econometrics. Econometric Society Publication No. 3. Cambridge University Press.

McDonald, John F. and Moffitt, Robert A. (1980) *The uses of Tobit analysis.* The Review of Economics and Statistics.

Olsen, Randall J. (1978) *Note on the uniqueness of the maximum likelihood estimator for the Tobit model.* Econometrica, Vol.46, No. 5.

Olsen, Randall J. (1980) *Notes and Comments. A Least Squares Correction for Selectivity Bias.* Econometrica, Vol. 48, No. 7.

Tobin, James (1958) *Estimation of relationships for limited dependent variables.* Econometrica, Vol.26, 24-36.

Wang, Liqun (1998) *Estimation of censored linear errors-in-variables models.* Journal of Econometrics, 84, 383-400.

Zellner, Arnold (1976) *Bayesian and Non-Bayesian Analysis of the Regression Model with Multivariate Student-t Error terms.* Journal of the American Statistical Association. Vol. 71, No. 354, Theory and Methods Section.