Econometrics II
Tutorial Problems No. 3

Lennart Hoogerheide & Agnieszka Borowska

01.03.2017

1 Summary

- **Limited dependent variable:** A continuous dependent variable which can take only a limited range of values (due to censoring or truncation).
- **Truncated Data Sample:** A sample from which some observations have been systematically excluded.

  [E.g. a sample of households with incomes under $200,000 explicitly excludes households with incomes over that level; thus: is not a **random sample of all** households.]
- **Censored Data Sample:** A sample from which no observations have been systematically excluded, but some of the information contained in them has been suppressed.

  [E.g. a sample of households in which all income levels are included, but for those with incomes in excess of $200,000, the amount reported is always exactly $200,000 (to protect the privacy of high-income respondents).]
- **BLUE estimator:** Best Linear Unbiased Estimator
  (the OLS estimator for the linear regression model under the Gauss-Markov assumptions, in particular: $\mathbb{E}(u|X) = 0$ and $\mathbb{E}(uu'|X) = \sigma^2 I$).
- **Truncated Regression Model:** A linear regression model for cross-sectional data in which the sampling scheme entirely excludes, on the basis of outcomes on the dependent variable, part of the population.
- **Truncated Normal Regression Model:** The special case of the truncated regression model where the underlying population model satisfies the classical linear model assumptions.
- **Probability mass function:** (pmf) a function that gives the probability that a discrete random variable is exactly equal to some value.
- **Probability density function:** (pdf) a function, whose value at any sample (or point) in the sample space can be interpreted as providing a *relative likelihood* that the value of the (continuous) random variable would equal that sample (because the *absolute likelihood* for a continuous random variable to take on any particular value is 0). The pdf is used to specify the probability of the random variable falling within a particular *range* of values (as opposed to taking on any one value).
- **Mixed probability distribution:** a probability distribution which is a mixture (i.e. a weighted sum) of different distributions (the weights correspond to the probabilities of different components occurring).

  [E.g. a mixed discrete/continuous distribution is 'partially' discrete and 'partially' continuous]
- **Censored Regression Model:** A multiple regression model where the dependent variable has been censored above and/or below some known threshold.
- **Censored Normal Regression Model:** The special case of the censored regression model where the underlying population model satisfies the classical linear model assumptions.
- **Tobit Model:** A censored normal regression model, with left-censoring at 0.
- **Corner Solution Response:** Censored data (so the same model for estimation is used) with different (truncated) interpretation: we are interested in the observed uncensored data themselves (so we want to know $E(y_i|x_i)$), while for censored data we are actually interested in the (partially unobserved) data "before censoring" (so we want to know $E(y_i^*|x_i)$).
- **Selected Sample:** A sample of data obtained not by random sampling but by selecting on the basis of some observed or unobserved characteristic.

2 Extra Topic: Prediction and marginal effects from the censored regression model[1]

There are potentially three conditional means of interest, and the resulting partial effects, in a censored regression model (in particular: in the Tobit model). Which one is the "right" one, depends on the research question or the purpose of study. We can consider:

⇒ the **index/latent** variable y^*:

$$\mathbb{E}(y_i^*|x_i) = x_i'\beta \quad \Longrightarrow \quad \frac{\partial \mathbb{E}(y_i^*|x_i)}{\partial x_i} = \beta$$

(which might be hard to interpret as y_i^* is unobserved);

⇒ the observed **censored** variable y, drawn from the whole population:

$$\mathbb{E}(y_i|x_i) = ?? \quad \Longrightarrow \quad \frac{\partial \mathbb{E}(y_i|x_i)}{\partial x_i} = ??$$

(usually used for predictions from the model);

⇒ the observed **uncensored** variable y, i.e. *conditionally* on $y^* > 0$, drawn from the (truncated) subpopulation

$$\mathbb{E}(y_i|y_i > 0, x_i) = ?? \quad \Longrightarrow \quad \frac{\partial \mathbb{E}(y_i|y_i > 0, x)}{\partial x} = ??.$$

We will derive the results for the second, censored case, and the results for the third, truncated case will follow. The goal is to derive

$$\mathbb{E}(y_i|x_i) = \Phi\left(\frac{x_i'\beta}{\sigma}\right) \cdot x_i'\beta + \sigma \cdot \phi\left(\frac{x_i'\beta}{\sigma}\right), \qquad (17.25)$$

(which we need for the computer exercise) and

$$\frac{\partial \mathbb{E}(y_i|x)}{\partial x} = \beta \cdot \Phi\left(\frac{x_i'\beta}{\sigma}\right).$$

The theorem below, together with the proof, are given for the general case of double sided censoring (the results for the Tobit model can be obtained as a special case).

[1] Based on Greene (2010), "Econometric Analysis", Chapter 19.

Theorem: Partial Effects in the Censored Regression Model

In the censored regression model with latent regression $y^* = x'\beta + \varepsilon$ and observed dependent variable

$$y_i = \begin{cases} a, & \text{if } y_i^* \leq a, \\ y_i^*, & \text{if } a < y_i^* < b, \\ b, & \text{if } y_i^* \geq b, \end{cases}$$

where a and b are constants, let $f(\varepsilon)$ and $F(\varepsilon)$ denote the density and cdf of ε. Assume that ε is a continuous random variable with mean 0 and variance σ^2, and $f(\varepsilon|x) = f(\varepsilon)$. Then:

$$\frac{\partial \mathbb{E}(y|x)}{\partial x} = \beta \cdot \mathbb{P}(y^* \in (a,b)).$$

Proof. By definition

$$\begin{aligned}
\mathbb{E}(y|x) &= a \cdot \mathbb{P}(y = a|x) + \mathbb{E}(y|y \in (a,b), x) \cdot \mathbb{P}(y \in (a,b)|x) + b \cdot \mathbb{P}(y = b|x) \\
&= a \cdot \mathbb{P}(y^* \leq a|x) + \mathbb{E}(y^*|y^* \in (a,b), x) \cdot \mathbb{P}(y^* \in (a,b)|x) + b \cdot \mathbb{P}(y^* \geq b|x) \\
&= a \cdot \mathbb{P}(x'\beta + \varepsilon \leq a|x) + \mathbb{E}(y^*|y^* \in (a,b), x) \cdot \mathbb{P}(a < x'\beta + \varepsilon < b|x) + b \cdot \mathbb{P}(x'\beta + \varepsilon \geq b|x) \\
&= a \cdot \mathbb{P}(\varepsilon \leq a - x'\beta|x) + \mathbb{E}(y^*|y^* \in (a,b), x) \cdot \mathbb{P}(a - x'\beta < \varepsilon < b - x'\beta|x) + b \cdot \mathbb{P}(\varepsilon \geq b - x'\beta|x) \\
&= a \cdot \mathbb{P}\left(\frac{\varepsilon}{\sigma} \leq \frac{a - x'\beta}{\sigma}\middle| x\right) + \mathbb{E}(y^*|y^* \in (a,b), x) \cdot \mathbb{P}\left(\frac{a - x'\beta}{\sigma} < \frac{\varepsilon}{\sigma} < \frac{b - x'\beta}{\sigma}\middle| x\right) + b \cdot \mathbb{P}\left(\frac{\varepsilon}{\sigma} \geq \frac{b - x'\beta}{\sigma}\middle| x\right).
\end{aligned} \tag{1}$$

Denote $z = \frac{\varepsilon}{\sigma}$,

$$\begin{aligned}
A &= \frac{a - x'\beta}{\sigma}, \qquad F_a = F(A), \qquad f_a = f(A), \\
B &= \frac{b - x'\beta}{\sigma}, \qquad F_b = F(B), \qquad f_b = f(B),
\end{aligned}$$

so that (1) becomes

$$\begin{aligned}
\mathbb{E}(y|x) &= a \cdot \mathbb{P}\left(\frac{\varepsilon}{\sigma} \leq \frac{a - x'\beta}{\sigma}\middle| x\right) + \mathbb{E}(y^*|y^* \in (a,b), x) \cdot \mathbb{P}\left(\frac{a - x'\beta}{\sigma} < \frac{\varepsilon}{\sigma} < \frac{b - x'\beta}{\sigma}\middle| x\right) + b \cdot \mathbb{P}\left(\frac{\varepsilon}{\sigma} \geq \frac{b - x'\beta}{\sigma}\middle| x\right) \\
&= a \cdot \mathbb{P}(z \leq A|\, x) + \mathbb{E}(y^*|y^* \in (a,b), x) \cdot \mathbb{P}(A < z < B|\, x) + b \cdot \mathbb{P}(z \geq B|\, x) \\
&= a \cdot F_a + \underbrace{\mathbb{E}(y^*|y^* \in (a,b), x)}_{(\star)} \cdot (F_b - F_a) + b \cdot (1 - F_b).
\end{aligned}$$

Next, we want to obtain the $(\star)$ term, i.e. the conditional mean of the continuous variable. Notice that this is the expectation of the truncated variable, $\mathbb{E}(y|y \in (a,b), x)$, i.e. expectation of y conditionally on y falling between the truncation points a and b. Hence, it will also answer our third question. By properties of the conditional expectation:

$$\begin{aligned}
\mathbb{E}(y^*|y^* \in (a,b), x) &= \mathbb{E}(x'\beta + \varepsilon|\, a < x'\beta + \varepsilon < b, x) \\
&= x'\beta + \mathbb{E}(\varepsilon|\, a - x'\beta < \varepsilon < b - x'\beta, x) \\
&= x'\beta + \sigma \mathbb{E}\left(\frac{\varepsilon}{\sigma}\middle| \frac{a - x'\beta}{\sigma} < \frac{\varepsilon}{\sigma} < \frac{b - x'\beta}{\sigma}, x\right) \\
&= x'\beta + \sigma \mathbb{E}(z|\, A < z < B, x) \\
&\overset{(*)}{=} x'\beta + \sigma \int_A^B \frac{z f(z)}{F_b - F_a} dz, \\
&= x'\beta + \frac{\sigma}{F_b - F_a} \int_A^B z f(z) dz,
\end{aligned} \tag{2}$$

where normalising by a constant $(F_b - F_a)$ in $(*)$ is due to truncation.

Collecting (1) and (2) gives us the desired expectation of the censored variable:

$$\begin{aligned}\mathbb{E}(y|x) &= a \cdot F_a + \mathbb{E}(y^*|y^* \in (a,b), x) \cdot (F_b - F_a) + b \cdot (1 - F_b) \\ &= a \cdot F_a + \left[x'\beta + \frac{\sigma}{F_b - F_a} \int_A^B z f(z) dz \right] \cdot (F_b - F_a) + b \cdot (1 - F_b) \\ &= a \cdot F_a + x'\beta \cdot (F_b - F_a) + \sigma \underbrace{\int_A^B z f(z) dz}_{(\blacksquare)} + b \cdot (1 - F_b). \end{aligned} \tag{3}$$

What is only left is to differentiate (3) wrt to x. Notice that differentiating of the cdf $F_\bullet$ wrt respect to x gives us the pdf $f_\bullet \cdot \left(\frac{-\beta}{\sigma}\right)$ ($\bullet = a, b$), where the last term obviously follows form the chain rule. Notice, that in ($\blacksquare$) the only place where x is present are the limits of integration. Hence, we need to use Leibnitz's integral rule[2] as follows:

$$\begin{aligned}\frac{\partial \mathbb{E}(y|x)}{\partial x} =& a \cdot f_a \cdot \left(\frac{-\beta}{\sigma}\right) + \beta \cdot (F_b - F_a) + x'\beta \cdot \left[f_b \cdot \left(\frac{-\beta}{\sigma}\right) - f_a \cdot \left(\frac{-\beta}{\sigma}\right)\right] + \frac{\partial}{\partial x} \sigma \int_A^B z f(z) dz - b \cdot f_b \cdot \left(\frac{-\beta}{\sigma}\right) \\ & \left\{ \frac{dA}{dt} = -\frac{\beta}{\sigma},\ z f(z)|_A = A f_a \right\} \\ =& a \cdot f_a \cdot \left(\frac{-\beta}{\sigma}\right) + \beta \cdot (F_b - F_a) + x'\beta \cdot \left[f_b \cdot \left(\frac{-\beta}{\sigma}\right) - f_a \cdot \left(\frac{-\beta}{\sigma}\right)\right] + \sigma \cdot (B f_b - A f_a) \cdot \left(-\frac{\beta}{\sigma}\right) - b \cdot f_b \cdot \left(\frac{-\beta}{\sigma}\right). \end{aligned}$$

Finally, we simplify by cancelling out terms in the above expression (using the definitions of A and B), to obtain:

$$\begin{aligned}\frac{\partial \mathbb{E}(y|x)}{\partial x} &= \beta \cdot (F_b - F_a) \\ &= \beta \cdot \mathbb{P}(y_i^* \in (a,b)).\end{aligned}$$

□

[2] Leibnitz's integral rule for differentiation under the integral sign states that:

$$\frac{d}{dt} \int_{a(t)}^{b(t)} f(x,t) dx = f(b(t), t) \cdot \frac{db(t)}{dt} - f(a(t), t) \cdot \frac{da(t)}{dt} + \int_{a(t)}^{b(t)} \frac{df(x,t)}{dt} dx,$$

where in our case the last term drops out because $f(z)$ does not depend on x.

The result from the theorem applied to the particular case of the original Tobit model (with left-censoring at 0) simplifies to[3]:

$$\frac{\partial \mathbb{E}(y_i|x_i)}{\partial x_i} = \beta \cdot \Phi\left(\frac{x_i'\beta}{\sigma}\right).$$

Roughly speaking, it suggests that the OLS estimates of the coefficients in a Tobit model usually resemble the MLEs times the proportion of nonlimit observations in the sample.

Hence, the marginal effects in the case of censoring are not β but smaller, with reduction factor $\Phi\left(\frac{x_i'\beta}{\sigma}\right)$:

- the difference will be small for large values of $\frac{x_i'\beta}{\sigma}$, as then $\Phi\left(\frac{x_i'\beta}{\sigma}\right) \approx 1$;
- the difference will be large for small values of $\frac{x_i'\beta}{\sigma}$, as then $\Phi\left(\frac{x_i'\beta}{\sigma}\right) \approx 0$.

The intuition should be clear: we observe a positive $y_i > 0$ when $y_i^* = x_i'\beta + \varepsilon_i > 0$, so the condition for observing an uncensored variable is $z_i = \frac{\varepsilon_i}{\sigma} > -\frac{x_i'\beta}{\sigma}$.

- If $\frac{x_i'\beta}{\sigma}$ is high and positive, then this is a non-restrictive condition and we will usually observe $y_i = y_i^*$. So when there is hardly any censoring, the marginal effects will be almost the same as in the standard regression model, i.e. β.
- If $\frac{x_i'\beta}{\sigma}$ is high and negative, then this is a very restrictive condition and we will usually observe the censored $y_i = 0$. So when there is a "hard" censoring, the marginal effects will be negligible, and only via an increase in the probability of recording a non-censored observation.

Hence, notice that the marginal effect of the explanatory variables in the Tobit model can be decomposed in two parts: when $x_i'\beta$ increases and

⇒ if $y_i = 0$, then the probability of $y_i > 0$ (a positive response) increases (i.e. the probability of falling in the positive part of the distribution);

⇒ if $y_i > 0$, then the mean response increases (i.e. the conditional mean of y^*).

3 Lecture Problems

Exercise 5

Suppose that we only started keeping track of these machine parts after 2 years and that by now all machine parts are broken. That is, we now have left-truncated data where we only observe $y_i^* > \ln(2)$ (instead of right-truncated data with $y_i^* < \ln(1) = 0$).

(a) *Derive the probability density function (pdf) of y_i in this case.*

Underlying population that satisfies all the classical linear model assumptions:

$$y_i^* = x_i'\beta + u_i, \qquad u_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$$

where each u_i is independent from each x_j $(i, j = 1, 2, \ldots, n)$.

Left-truncated variable y_i:

$$y_i = \begin{cases} \text{not observed}, & \text{if } y_i^* \leq \ln(2), \\ y_i^*, & \text{if } y_i^* > \ln(2). \end{cases}$$

Here: boundary $c = \ln(2)$ for log-durations.

[3] Please check! Notice that then $a = 0$, there is no b (or, formally, $b = \infty$) and $F = \Phi$ and $f = \phi$.

We start with deriving the **cumulative distribution function** (CDF) of the truncated observation y_i (given x_i)[4], which is equal to the conditional probability $\mathbb{P}(y_i^* \leq a | y_i^* > c)$ for $a > c$:

$$\begin{aligned}
\mathbb{P}(y_i \leq a) = \mathbb{P}(y_i^* \leq a | y_i^* > c)& \\
&\overset{(*)}{=} \mathbb{P}(y_i^* \leq a \text{ and } y_i^* > c | y_i^* > c) \\
&= \frac{\mathbb{P}(c < y_i^* \leq a)}{\mathbb{P}(y_i^* > c)} \\
&\overset{(**)}{=} \frac{\mathbb{P}\left(\frac{c - x_i'\beta}{\sigma} < \frac{y_i^* - x_i'\beta}{\sigma} \leq \frac{a - x_i'\beta}{\sigma}\right)}{\mathbb{P}\left(\frac{y_i^* - x_i'\beta}{\sigma} > \frac{c - x_i'\beta}{\sigma}\right)} \\
&= \frac{\Phi\left(\frac{a - x_i'\beta}{\sigma}\right) - \Phi\left(\frac{c - x_i'\beta}{\sigma}\right)}{1 - \Phi\left(\frac{c - x_i'\beta}{\sigma}\right)},
\end{aligned}$$

where in $(*)$ we used that $c < a$ (so that $y_i^* \leq a$ and $y_i^* > c$ imply $c < y_i^* \leq a$) and in $(**)$ that $\frac{y_i^* - x_i'\beta}{\sigma}$ has standard normal distribution $\mathcal{N}(0, 1)$.

Then, the **probability density function** (pdf) of y_i is given by the derivative of the cdf:

$$\begin{aligned}
p_{y_i}(a) &= \frac{\partial \mathbb{P}(y_i \leq a)}{\partial a} \\
&= \frac{\partial \Phi\left(\frac{a - x_i'\beta}{\sigma}\right)}{\partial a} \cdot \frac{1}{1 - \Phi\left(\frac{c - x_i'\beta}{\sigma}\right)} \\
&= \frac{\frac{1}{\sigma}\phi\left(\frac{a - x_i'\beta}{\sigma}\right)}{1 - \Phi\left(\frac{c - x_i'\beta}{\sigma}\right)}.
\end{aligned}$$

(b) *Derive the log-likelihood* $\ln L(\beta, \sigma)$.

The likelihood function (of the whole sample) is given by:

$$\begin{aligned}
L(\beta, \sigma) &= p(y_1, \ldots, y_n | x_1, \ldots, x_n) \\
&\overset{(*)}{=} \prod_{i=1}^{n} p(y_i | x_i) \\
&= \prod_{i=1}^{n} \frac{\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)}{1 - \Phi\left(\frac{c - x_i'\beta}{\sigma}\right)},
\end{aligned}$$

where $(*)$ holds because $y_1, \ldots, y_n$ are independent (conditionally upon $x_1, \ldots, x_n$).

And the loglikelihood is simply the logarithm of the likelihood:

$$\begin{aligned}
\ln L(\beta, \sigma) &= \sum_{i=1}^{n} \ln p(y_i | x_i) \\
&= \sum_{i=1}^{n} \left\{ -\ln(\sigma) + \ln\left[\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right] - \ln\left[1 - \Phi\left(\frac{c - x_i'\beta}{\sigma}\right)\right] \right\}.
\end{aligned}$$

Note that maximizing $\ln L(\beta, \sigma)$ (using numerical optimization method like BFGS) yields $\hat{\beta}_{ML}$ and $\hat{\sigma}_{ML}$.

[4] Note: all probabilities below are conditional upon x_i (dropped from notation to make formulas (hopefully) clearer).

Exercise 6

Derive the log-likelihood in a linear regression model where the dependent variable is left-truncated (with bound 0) ***and*** *right-censored (with bound 1). That is:*

$$\begin{aligned} y_i^* &= x_i'\beta + u_i, \\ u_i &\sim \mathcal{N}(0, \sigma^2), \\ y_i &= \begin{cases} \textit{not observed}, & \textit{if } y_i^* \leq 0, \\ y_i^*, & \textit{if } 0 < y_i^* < 1, \\ 1, & \textit{if } y_i^* \geq 1. \end{cases} \end{aligned}$$

First derive the probability $\mathbb{P}(y_i = 1|x_i)$ and the density for y_i (for $0 < y_i < 1$).

The probability $\mathbb{P}(y_i = 1|x_i)$ is the conditional probability $\mathbb{P}(y_i^* \geq 1|y_i^* > 0)$, because we only observe observations with $y_i^* > 0$ (where the conditioning upon x_i is again dropped from the notation):

$$\begin{aligned} \mathbb{P}(y_i^* \geq 1|y_i^* > 0) &= \frac{\mathbb{P}(y_i^* \geq 1)}{\mathbb{P}(y_i^* > 0)} \\ &= \frac{\mathbb{P}\left(x_i'\beta + u_i \geq 1\right)}{\mathbb{P}\left(x_i'\beta + u_i > 0\right)} \\ &= \frac{\mathbb{P}(u_i \geq 1 - x_i'\beta)}{\mathbb{P}(u_i > 0 - x_i'\beta)} \\ &= \frac{\mathbb{P}\left(\frac{u_i}{\sigma} \geq \frac{1-x_i'\beta}{\sigma}\right)}{\mathbb{P}\left(\frac{u_i}{\sigma} > \frac{0-x_i'\beta}{\sigma}\right)} \\ &= \frac{1 - \mathbb{P}\left(\frac{u_i}{\sigma} < \frac{1-x_i'\beta}{\sigma}\right)}{1 - \mathbb{P}\left(\frac{u_i}{\sigma} \leq \frac{0-x_i'\beta}{\sigma}\right)} \\ &= \frac{1 - \Phi\left(\frac{1-x_i'\beta}{\sigma}\right)}{1 - \Phi\left(\frac{0-x_i'\beta}{\sigma}\right)}, \end{aligned}$$

where we used that $\frac{u_i}{\sigma}$ has a standard normal distribution.

The density for y_i (for $0 < y_i < 1$) is the density in the left-truncated model (with boundary $c = 0$). From Exercise 5 we already have the pdf:

$$\begin{aligned} p_{y_i}(a) &= \frac{\frac{1}{\sigma}\phi\left(\frac{a-x_i'\beta}{\sigma}\right)}{1 - \Phi\left(\frac{c-x_i'\beta}{\sigma}\right)} \\ &= \frac{\frac{1}{\sigma}\phi\left(\frac{a-x_i'\beta}{\sigma}\right)}{1 - \Phi\left(\frac{0-x_i'\beta}{\sigma}\right)}. \end{aligned}$$

Note: censoring does **not** affect the pdf of those observations that are not censored. Whereas truncation does affect the pdf of those observations that are not truncated.

Likelihood: product of probability density functions (♠) (for $y_i < 1$ with continuous distribution) and probability functions (♣) (for $y_i = 1$ with discrete distribution) with observed y_i (and x_i) substituted:

$$\begin{aligned} L(\beta, \sigma) &= p(y_1, \ldots, y_n|x_1, \ldots, y_n) \\ &\overset{(*)}{=} \prod_{i=1}^{n} p(y_i|x_i) \\ &= \underbrace{\prod_{\{y_i<1\}} \left[\frac{\frac{1}{\sigma}\phi\left(\frac{y_i-x_i'\beta}{\sigma}\right)}{1 - \Phi\left(\frac{0-x_i'\beta}{\sigma}\right)}\right]}_{(\spadesuit)} \times \underbrace{\prod_{\{y_i=1\}} \left[\frac{1 - \Phi\left(\frac{1-x_i'\beta}{\sigma}\right)}{1 - \Phi\left(\frac{0-x_i'\beta}{\sigma}\right)}\right]}_{(\clubsuit)}, \end{aligned}$$

where $(*)$ holds because $y_1, \dots, y_n$ are independent (conditionally upon $x_1, \dots, x_n$).

Then, the loglikelihood is:

$$\ln L(\beta, \sigma) = \sum_{i=1}^{n} \ln p(y_i|x_i) =$$
$$= \underbrace{\sum_{\{y_i<1\}} \left\{ -\ln(\sigma) + \ln\left[\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right] - \ln\left[1 - \Phi\left(\frac{0 - x_i'\beta}{\sigma}\right)\right]\right\}}_{(\spadesuit)}$$
$$+ \underbrace{\sum_{\{y_i=1\}} \left\{ \ln\left[1 - \Phi\left(\frac{1 - x_i'\beta}{\sigma}\right)\right] - \ln\left[1 - \Phi\left(\frac{0 - x_i'\beta}{\sigma}\right)\right]\right\}}_{(\clubsuit)}.$$

Note: maximizing $\ln L(\beta, \sigma)$ (using numerical optimization method like BFGS) yields $\hat{\beta}_{ML}$ and $\hat{\sigma}_{ML}$.

4 Exercises

4.1 W17/6

Consider a family saving function for the population of all families in the United States:

$$sav = \beta_0 + \beta_1 inc + \beta_2 hhsize + \beta_3 educ + \beta_4 age + u,$$

where hhsize is household size, educ is years of education of the household head, and age is age of the household head. Assume that $\mathbb{E}(u|inc, hhsize, educ, age) = 0$.

(i) *Suppose that the sample includes only families whose head is over* 25 *years old. If we use OLS on such a sample, do we get unbiased estimators of the* β_j*? Explain.*

   OLS will be unbiased, because we are choosing the sample on the basis of an exogenous explanatory variable. The population regression function for *sav* is the same as the regression function in the subpopulation with $age > 25$.

(ii) *Now, suppose our sample includes only married couples without children. Can we estimate all of the parameters in the saving equation? Which ones can we estimate?*

   Assuming that marital status and number of children affect *sav* only through household size (*hhsize*), this is another example of exogenous sample selection. But, in the subpopulation of married people without children, $hhsize = 2$. Because there is no variation in *hhsize* in the subpopulation, we would not be able to estimate β_2. Effectively, the intercept in the subpopulation becomes $\beta_0 + 2\beta_2$, and that is all we can estimate. But, assuming there is variation in *inc*, *educ*, and *age* among married people without children (and that we have a sufficiently varied sample from this subpopulation), we can still estimate β_1, β_3 and β_4.

(iii) *Suppose we exclude from our sample families that save more than $25,000 per year. Does OLS produce consistent estimators of the* β_j*?*

   This would be selecting the sample on the basis of the dependent variable, which causes OLS to be biased and inconsistent for estimating the β in the population model. We should instead use a truncated regression model.

4.2 Double censoring problem

Management consultants working for a very large consultancy firm AwesomeConsulting are assigned to a number of projects depending on their characteristics, collected in a $k \times 1$ *vector* x_i' *for individual* i *(including their salary, experience, etc.). We want to model their weekly chargeable hours* y_i*. We have a random sample of* N *independent observations on* y_i *and corresponding* x_i'*. For simplicity we model the regular number of hours as a continuous variable, but take into account the possibility that during a week there might be no chargeable hours and that the maximum number of hours that can be charged to a client is by contract limited to* 40 *hours.*

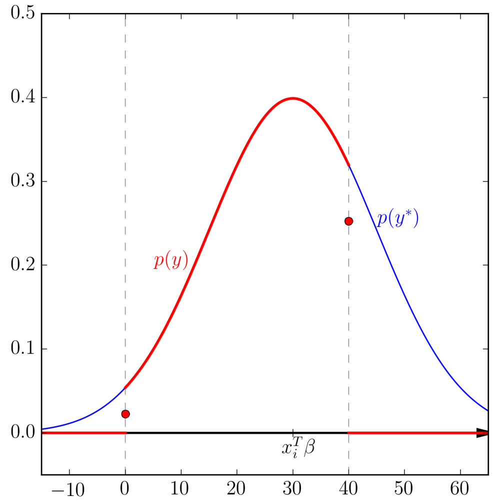


Figure 1: Double censoring: left censoring at 0 and right censoring at 40. Example with the mean $x_i'\beta$ at 30 and the standard deviation $\sigma = 15$. Then $\mathbb{P}(y_i = 0|x_i) = \Phi\left(-\frac{x_i'\beta}{\sigma}\right) = 0.0228$, $\mathbb{P}(y_i = 40|x_i) = 1 - \Phi\left(\frac{40-x_i'\beta}{\sigma}\right) = 0.25258$ and $\mathbb{P}(0 < y_i < 40|x_i) = \int_0^{40} \phi(z)dz = 0.7247$.

(a) *Model this situation using a latent variable y^* given by:*

$$\begin{aligned} y_i^* &= x_i'\beta + u_i, \\ u_i &\overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2). \end{aligned}$$

Give the appropriate probability mass- and density functions for the different outcomes of the observed charged hours y. Give an interpretation and illustrate the situation graphically.

Standard censored regression model with left and right censoring (at 0 and 40) is given by:

$$\begin{aligned} y_i^* &= x_i'\beta + u_i, \\ u_i &\overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2), \\ y_i &= \begin{cases} 0, & \text{if } y_i^* \leq 0, \\ y_i^*, & \text{if } 0 < y_i^* < 40, \\ 40, & \text{if } y_i^* \geq 40. \end{cases} \end{aligned}$$

The probability mass functions at the censored value of 0 is the probability of *observing* the value of 0:

$$\begin{aligned} \mathbb{P}(y_i = 0|x_i) &= \mathbb{P}(y_i^* \leq 0|x_i) \\ &= \mathbb{P}(x_i'\beta + u_i \leq 0|x_i) \\ &= \mathbb{P}(u_i \leq -x_i'\beta|x_i) \\ &\overset{(*)}{=} \mathbb{P}\left(\frac{u_i}{\sigma} \leq -\frac{x_i'\beta}{\sigma} \middle| x_i\right) \\ &\overset{(**)}{=} \mathbb{P}\left(\frac{u_i}{\sigma} \leq -\frac{x_i'\beta}{\sigma}\right) \\ &= \Phi\left(-\frac{x_i'\beta}{\sigma}\right), \end{aligned}$$

where in $(*)$ we standardise u_i by dividing it by its standard deviation σ and in $(**)$ we use the assumption about independence of u_i and x_i.

Similarly, the probability mass functions at the censored value of 40 is the probability of *observing* the value of 40:

$$
\begin{aligned}
\mathbb{P}(y_i = 40|x_i) &= \mathbb{P}(y_i^* \geq 40|x_i) \\
&= \mathbb{P}(x_i'\beta + u_i \geq 40|x_i) \\
&= \mathbb{P}(u_i \geq 40 - x_i'\beta|x_i) \\
&\overset{(*)}{=} \mathbb{P}\left(\frac{u_i}{\sigma} \geq \frac{40 - x_i'\beta}{\sigma}\middle| x_i\right) \\
&\overset{(**)}{=} \mathbb{P}\left(\frac{u_i}{\sigma} \leq \frac{x_i'\beta - 40}{\sigma}\middle| x_i\right) \\
&\overset{(***)}{=} \mathbb{P}\left(\frac{u_i}{\sigma} \leq \frac{x_i'\beta - 40}{\sigma}\right) \\
&= \Phi\left(\frac{x_i'\beta - 40}{\sigma}\right) \\
&= \Phi\left(-\frac{40 - x_i'\beta}{\sigma}\right) \\
&\overset{(****)}{=} 1 - \Phi\left(\frac{40 - x_i'\beta}{\sigma}\right),
\end{aligned}
$$

where in $(*)$ we standardise u_i by dividing it by its standard deviation σ, in $(**)$ we use the symmetry of the standard normal distribution, in $(***)$ we use the assumption about independence of u_i and x_i and in $(****)$ the property of Φ, the CDF of the standard normal distribution: $\Phi(-x) = 1 - \Phi(x)$.

For continuous $y_i \in (0, 40)$ we use the probability density function. Because then

$$y_i = y_i^* = x_i'\beta + u_i,$$

with $u_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$, we have the standardised normal variable $\frac{u_i}{\sigma} = \frac{y_i - x_i'\beta}{\sigma}$ for which

$$p(y_i|x_i) = \frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right).$$

(b) *Derive the appropriate log-likelihood function for N independent observations.*

Now the likelihood is a product of probability density functions (♠) (for $0 < y_i < 40$ with continuous distribution) and *two* probability functions for y_i with discrete distributions: (♣) for $y_i = 40$ and (♡) for $y_i = 0$, with observed y_i (and x_i) substituted:

$$
\begin{aligned}
L(\beta, \sigma) =& p(y_1, \ldots, y_n|x_1, \ldots, y_n) \\
\overset{(*)}{=}& \prod_{i=1}^{n} p(y_i|x_i) \\
=& \underbrace{\prod_{\{0<y_i<40\}} \left[\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right]}_{(\spadesuit)} \times \underbrace{\prod_{\{y_i=40\}} \left[1 - \Phi\left(\frac{40 - x_i'\beta}{\sigma}\right)\right]}_{(\clubsuit)} \times \underbrace{\prod_{\{y_i=0\}} \Phi\left(-\frac{x_i'\beta}{\sigma}\right)}_{(\heartsuit)},
\end{aligned}
$$

where $(*)$ holds because $y_1, \ldots, y_n$ are independent (conditionally upon $x_1, \ldots, x_n$).

Then, the loglikelihood is:

$$
\begin{aligned}
\ln L(\beta, \sigma) =& \sum_{i=1}^{n} \ln p(y_i|x_i) = \\
&= \underbrace{\sum_{\{0<y_i<40\}} \left\{-\ln(\sigma) + \ln\left[\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right]\right\}}_{(\spadesuit)} \\
&+ \underbrace{\sum_{\{y_i=40\}} \left\{\ln\left[1 - \Phi\left(\frac{40 - x_i'\beta}{\sigma}\right)\right]\right\}}_{(\clubsuit)} + \underbrace{\sum_{\{y_i=0\}} \left\{\ln\Phi\left(-\frac{x_i'\beta}{\sigma}\right)\right\}}_{(\heartsuit)}.
\end{aligned}
$$

(c) *What is the marginal effect of salary (2nd element in x_i) on the possibility of individual i being fully (40 hours) chargeable?*

We need to differentiate the probability of being charged 40 hours with respect to the second variable, salary. We have:

$$\frac{\partial \mathbb{P}(y_i = 40)}{\partial x_{i2}} = \frac{\partial \mathbb{P}(y_i^* \geq 40)}{\partial x_{i2}} = \phi\left(\frac{40 - x_i'\beta}{\sigma}\right)\frac{\beta_2}{\sigma}.$$

Note that is it positive when $\beta_2 > 0$.

(d) *What problems in modelling can you expect in the following cases? Think about the validity of the model assumptions.*

(i) *The sample consists of a sample based on direct colleagues from the same branch.*
Contemporaneous correlation – causes observations to be non i.i.d..

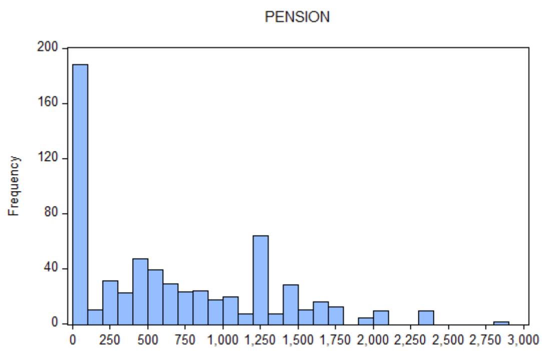
(ii) *The sample consists of a sample based on weeks for one individual such that i refers to the weeks in the sample?*
Serial correlation – causes observations to be non i.i.d..

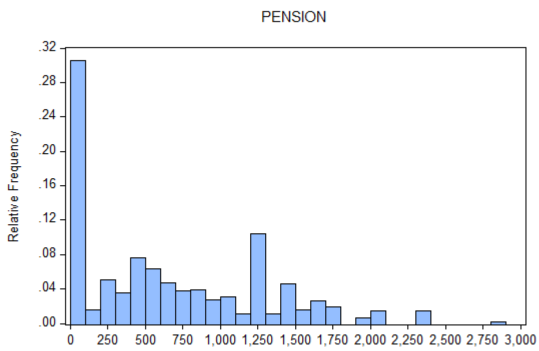
5 Computer Exercise

W17/C3

Use the data in `fringe.wf1` *for this exercise*[5].

(i) *For what percentage of the workers in the sample is pension equal to zero? What is the range of pension for workers with nonzero pension benefits? Why is a Tobit model appropriate for modelling pension?*





We can see that out of 616 workers, 172, or about 0.28%, have zero pension benefits. For the 444 workers reporting positive pension benefits, the range is from 7.28 to 2,880.27[6]. Therefore, we have a nontrivial fraction of the sample with $pension_i = 0$, and the range of positive pension benefits is fairly wide. The Tobit model is well-suited to this kind of dependent variable.

(ii) *Estimate a Tobit model explaining pension in terms of exper, age, tenure, educ, depends, married, white, and male. Do whites and males have statistically significant higher expected pension benefits?*

[5] $N = 616$, cross-sectional family data on pension benefits.
[6] You can easily check it in EViews by sorting *pension*: `pension.sort`.

Equation: EQ_TOBIT Workfile: FRINGE::Fringe\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: PENSION
Method: ML - Censored Normal (TOBIT) (Newton-Raphson / Marquardt steps)
Sample: 1 616
Included observations: 616
Left censoring (value) series: 0
Convergence achieved after 5 iterations
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-1252.429	219.0781	-5.716815	0.0000
EXPER	5.203458	6.009515	0.865870	0.3866
AGE	-4.638944	5.710965	-0.812287	0.4166
TENURE	36.02385	4.564528	7.892130	0.0000
EDUC	93.21262	10.89176	8.558083	0.0000
DEPENDS	35.28461	21.91775	1.609864	0.1074
MARRIED	53.68858	71.73541	0.748425	0.4542
WHITE	144.0855	102.0792	1.411507	0.1581
MALE	308.1505	69.89298	4.408890	0.0000
Error Distribution				
SCALE:C(10)	677.7383	24.14034	28.07493	0.0000

Mean dependent var	652.3368	S.D. dependent var	619.1199
S.E. of regression	532.1477	Akaike info criterion	11.95767
Sum squared resid	1.72E+08	Schwarz criterion	12.02948
Log likelihood	-3672.964	Hannan-Quinn criter.	11.98559
Avg. log likelihood	-5.962603		

Left censored obs	172	Right censored obs	0
Uncensored obs	444	Total obs	616

Being white or male (or, of course, both) increases predicted pension benefits, although only *male* is statistically significant with the z statistics (asymptotically equal to the t statistics) $z \approx 4.41$ and the corresponding p-value (i.e. `Prob.` in the EViews output) of 0.0000. For *white* the p-value of 0.1581 does not allow us to reject the null that its coefficient is equal 0 (at the standard significance level $\alpha = 0.05$).

(iii) *Use the results from part (ii) to estimate the difference in expected pension benefits for a white male and a nonwhite female, both of whom are* 35 *years old, are single with no dependence, have* 16 *years of education, and have* 10 *years of experience*[7].

We need to use formula (17.25) from the book, which is

$$\mathbb{E}(y|x) = \Phi\left(\frac{x^T\beta}{\sigma}\right)\cdot x^T\beta + \sigma\cdot\phi\left(\frac{x^T\beta}{\sigma}\right), \tag{17.25}$$

and describes the expected value of the dependent variable y in the Tobit model.

First, we consider $x^{(m)}$ with $white = 1$, $male = 1$, $age = 35$, $maried = 0$, $depends = 0$, $educ = 16$ and $exper = tenure = 10$. The linear index $x^{(m)T}\hat{\beta}$ is equal to

$$\begin{aligned} x^{(m)T}\hat{\beta} &= -1252.43 + 5.20\cdot 10 - 4.64\cdot 35 + 36.02\cdot 10 + 93.21\cdot 16 + 35.28\cdot 0 + 53.69\cdot 0 + 144.09\cdot 1 + 308.15\cdot 1 \\ &= 940.97. \end{aligned}$$

Second, we consider $x^{(f)}$ with $white = 0$, $male = 0$, $age = 35$, $maried = 0$, $depends = 0$, $educ = 16$ and $exper = tenure = 10$. The linear index $x^{(f)T}\hat{\beta}$ is equal to

$$\begin{aligned} x^{(f)T}\hat{\beta} &= -1252.43 + 5.20\cdot 10 - 4.64\cdot 35 + 36.02\cdot 10 + 93.21\cdot 16 + 35.28\cdot 0 + 53.69\cdot 0 + 144.09\cdot 0 + 308.15\cdot 0 \\ &= 488.73. \end{aligned}$$

Since the estimated standard deviation σ of the error term u_i is equal to $\hat{\sigma} = 677.74$ (c.f. `SCALE: C(10)`), we have

$$\begin{aligned} \mathbb{E}(pension|x^{(m)}) &= \Phi\left(\frac{x^{(m)T}\hat{\beta}}{\hat{\sigma}}\right)\cdot x^{(m)T}\hat{\beta} + \hat{\sigma}\cdot\phi\left(\frac{x^{(m)T}\hat{\beta}}{\hat{\sigma}}\right) \\ &= \Phi\left(\frac{940.97}{677.74}\right)\cdot 940.97 + 677.74\cdot\phi\left(\frac{940.97}{677.74}\right) \\ &= 0.92\cdot 940.97 + 677.74\cdot 0.15 \\ &= 966.49 \end{aligned}$$

[7] *Hint: use the formula (17.25) from the book for the expectation of the censored variable (in other words, for the predicted value from the Tobit model):*

$$\mathbb{E}(y|x) = \Phi\left(\frac{x^T\beta}{\sigma}\right)\cdot x^T\beta + \sigma\cdot\phi\left(\frac{x^T\beta}{\sigma}\right). \tag{17.25}$$

and

$$\begin{aligned}\mathbb{E}(pension|x^{(f)}) &= \Phi\left(\frac{x^{(f)T}\hat{\beta}}{\hat{\sigma}}\right)\cdot x^{(f)T}\hat{\beta} + \hat{\sigma}\cdot\phi\left(\frac{x^{(f)T}\hat{\beta}}{\hat{\sigma}}\right)\\ &= \Phi\left(\frac{488.73}{677.74}\right)\cdot 488.73 + 677.74\cdot\phi\left(\frac{488.73}{677.74}\right)\\ &= 0.76\cdot 488.73 + 677.74\cdot 0.31\\ &= 582.16,\end{aligned}$$

respectively. The difference in the expected pension value for a white male and for a nonwhite female with the same all other characteristics is thus

$$966.49 - 582.16 = 384.33.$$

(iv) Add union to the Tobit model and comment on its significance.

Equation: EQ_TOBIT2 Workfile: FRINGE::Fringe\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: PENSION
Method: ML - Censored Normal (TOBIT) (Newton-Raphson / Marquardt steps)
Sample: 1 616
Included observations: 616
Left censoring (value) series: 0
Convergence achieved after 5 iterations
Coefficient covariance computed using observed Hessian

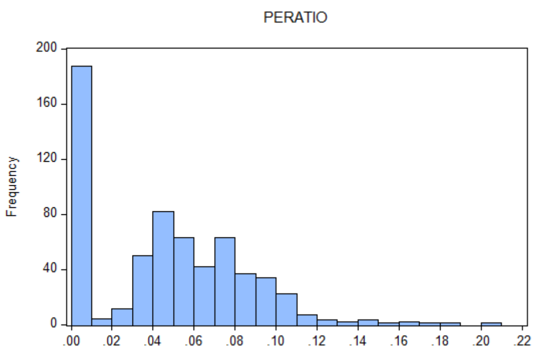
Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-1571.506	218.5445	-7.190784	0.0000
EXPER	4.393524	5.830947	0.753484	0.4512
AGE	-1.653532	5.555709	-0.297628	0.7660
TENURE	28.77837	4.504963	6.388147	0.0000
EDUC	106.8277	10.77274	9.916481	0.0000
DEPENDS	41.46623	21.21414	1.954650	0.0506
MARRIED	19.74555	69.50048	0.284107	0.7763
WHITE	159.2972	98.96748	1.609592	0.1075
MALE	257.2457	68.02052	3.781883	0.0002
UNION	439.0460	62.48832	7.026049	0.0000
Error Distribution				
SCALE:C(11)	652.8974	23.16287	28.18724	0.0000

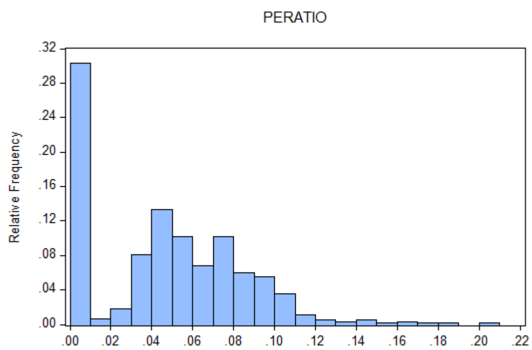
Mean dependent var	652.3368	S.D. dependent var	619.1199
S.E. of regression	518.5418	Akaike info criterion	11.88166
Sum squared resid	1.63E+08	Schwarz criterion	11.96065
Log likelihood	-3648.552	Hannan-Quinn criter.	11.91237
Avg. log likelihood	-5.922973		
Left censored obs	172	Right censored obs	0
Uncensored obs	444	Total obs	616

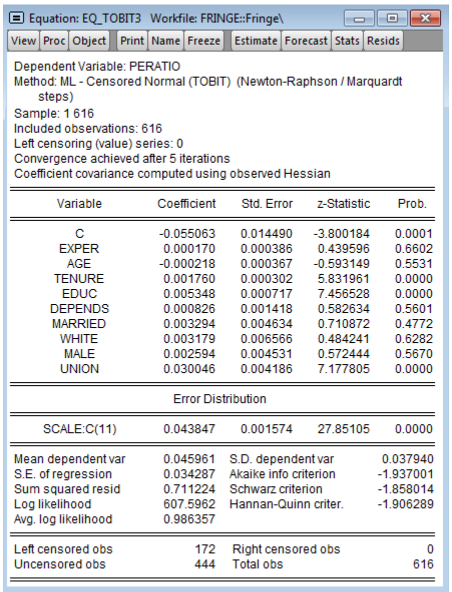
The estimated coefficient for *union* is 'large' (equal to 439.05) and significant (p-value=0.0000).

(v) Apply the Tobit model from part (iv) but with peratio, the pension-earnings ratio, as the dependent variable. (Notice that this is a fraction between zero and one, but, though it often takes on the value zero, it never gets close to being unity. Thus, a Tobit model is fine as an approximation.) Does gender or race have an effect on the pension-earnings ratio?

Indeed, the maximum value of *peratio* is less than 0.21, so a model with the right censoring is not needed.







Equation: EQ_TOBIT3 Workfile: FRINGE::Fringe\
View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: PERATIO
Method: ML - Censored Normal (TOBIT) (Newton-Raphson / Marquardt steps)
Sample: 1 616
Included observations: 616
Left censoring (value) series: 0
Convergence achieved after 5 iterations
Coefficient covariance computed using observed Hessian

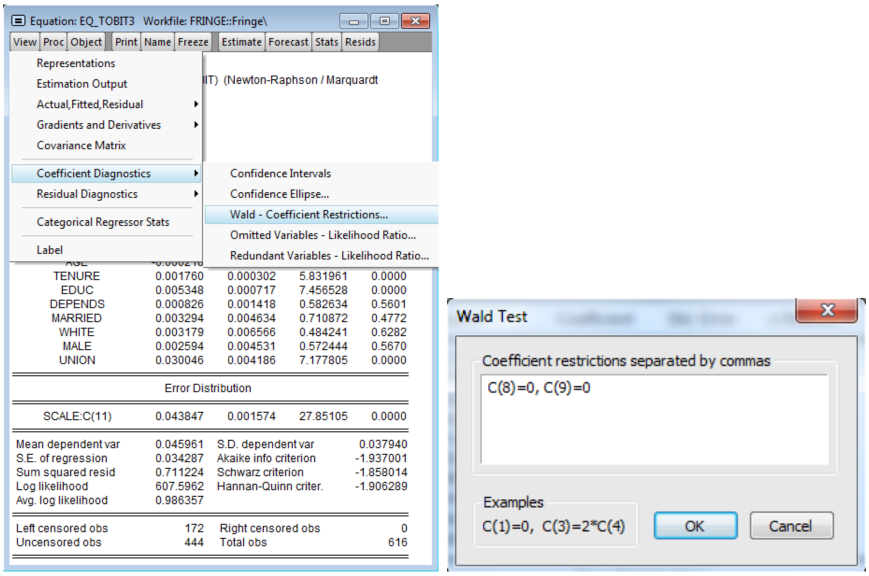
Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.055063	0.014490	-3.800184	0.0001
EXPER	0.000170	0.000386	0.439596	0.6602
AGE	-0.000218	0.000367	-0.593149	0.5531
TENURE	0.001760	0.000302	5.831961	0.0000
EDUC	0.005348	0.000717	7.456528	0.0000
DEPENDS	0.000826	0.001418	0.582634	0.5601
MARRIED	0.003294	0.004634	0.710872	0.4772
WHITE	0.003179	0.006566	0.484241	0.6282
MALE	0.002594	0.004531	0.572444	0.5670
UNION	0.030046	0.004186	7.177805	0.0000
	Error Distribution			
SCALE:C(11)	0.043847	0.001574	27.85105	0.0000

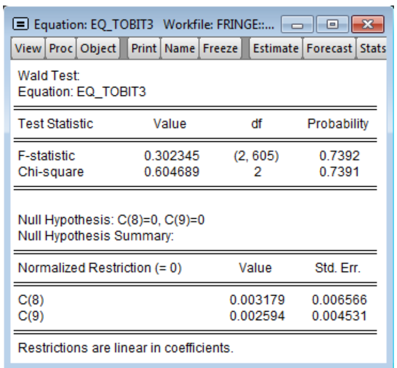
Mean dependent var	0.045961	S.D. dependent var	0.037940
S.E. of regression	0.034287	Akaike info criterion	-1.937001
Sum squared resid	0.711224	Schwarz criterion	-1.858014
Log likelihood	607.5962	Hannan-Quinn criter.	-1.906289
Avg. log likelihood	0.986357		

Left censored obs	172	Right censored obs	0
Uncensored obs	444	Total obs	616

When *peratio* is used as the dependent variable in the Tobit model, both *white* and *male* become insignificant (with the p-values of 0.6282 and 0.5670, respectively).

We can also check the joint significance of these two variables. For that, we can run the Wald test as shown below.





Equation: EQ_TOBIT3 Workfile: FRINGE::...
View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats

Wald Test:
Equation: EQ_TOBIT3

Test Statistic	Value	df	Probability
F-statistic	0.302345	(2, 605)	0.7392
Chi-square	0.604689	2	0.7391

Null Hypothesis: C(8)=0, C(9)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(8)	0.003179	0.006566
C(9)	0.002594	0.004531

Restrictions are linear in coefficients.

The resulting F statistic is equal to 0.30 with the corresponding p-value of 0.7392. So at any reasonable significance level we cannot reject the null that jointly *white* and *male* are insignificant.

Therefore, neither whites nor males seem to have different preferences for pension benefits as a fraction of earnings. White males have higher pension benefits because they have, on average, higher earnings.