UNIVERSITY OF LISBON ISEG
2020/2021 PAULO M.D.C. PARENTE

Advanced Econometrics
PhD in Economics
Exercise Sheet 4 - Limited Dependent Variable Models

1. Let t_i^* denote the duration of some event, such as unemployment, measured in continuous time. Consider the following model for t_i^*:

$$\begin{aligned} t_i^* &= \exp\left(\mathbf{X}_i'\boldsymbol{\beta}_0 + u_i\right), \; u_i|X_i \sim N(0,\sigma^2), \\ t_i &= \min\{t_i^*, c\} \\ &= \begin{cases} t_i^* & , \text{ if } t_i^* \leq c \\ c & , \text{ if } t_i^* > c \end{cases}, \end{aligned}$$

where $c > 0$ is a known censoring constant.

   (a) Find $P(t_i = c|X_i)$, that is, the probability that the duration is censored. What happens as $c \to \infty$?

   (b) What is the density of $\log(t_i)$ (given $\mathbf{X}_i$) when $t_i < c$? Now write down the full density of $\log(t_i)$ (given $\mathbf{X}_i$).

   (c) Write down the log-likelihood function for observation i.

   (d) Partition $\boldsymbol{\beta}_0$ into the K_1 and K_2 sub-vectors $\boldsymbol{\beta}_{0,1}$ and $\boldsymbol{\beta}_{0,2}$. Explain briefly how to test H_0: $\boldsymbol{\beta}_{0,2} = 0$.

   (e) Obtain the log-likelihood function if the censoring time is potentially different for each person, so that $t_i = \min\{t_i^*, c_i\}$, where c_i is observed for all i. Assume that u_i is independent of $(\mathbf{X}_i, c_i)$.

2. Suppose that, for a random draw (Y_i, X_i) from the population, Y_i is a doubly censored variable:

$$\begin{aligned} Y_i &= \min\{\max\{a_1, Y_i^*\}, a_2\} \\ &= \begin{cases} a_1 & Y_i^* < a_1 \\ Y_i^* & a_1 \leq Y_i^* \leq a_2 \\ a_2 & Y_i^* > a_2 \end{cases}, \end{aligned}$$

where $a_1 < a_2$ and $Y_i^*|X_i \sim N(\mathbf{X}_i'\boldsymbol{\beta}_0, \sigma^2)$.

   (a) Find $P(Y_i = a_1|\mathbf{X}_i)$ and $P(Y_i = a_2|\mathbf{X}_i)$ in terms of the standard normal cumulative distribution function, $\mathbf{X}_i$, $\boldsymbol{\beta}_0$, and σ.

   (b) For $y_i \in (a_1, a_2)$, find $P(Y_i \leq y_i|\mathbf{X}_i)$ and use this to find the density function of Y_i given $\mathbf{X}_i$.

   (c) If $z \sim N(0,1)$ it can be shown that

$$E\left(z|c_1 < z < c_2\right) = \frac{\phi\left(c_1\right) - \phi\left(c_2\right)}{\Phi\left(c_1\right) - \Phi\left(c_2\right)}$$

   for $c_1 < c_2$. Use this fact to find $E\left(Y_i|a_1 < Y_i < a_2, \mathbf{X}_i\right)$ and $E\left(Y_i|\mathbf{X}_i\right)$

(d) Consider the following method for estimating $\boldsymbol{\beta}_0$. Using only the uncensored observations, that is, observations for which $a_1 < Y_i < a_2$, run the OLS regression of Y_i on $\mathbf{X}_i$. Explain why this does not generally produce a consistent estimator of $\boldsymbol{\beta}_0$.

(e) Write down the log-likelihood function for observation i; it should consist of three parts.

(f) For data censoring, how would the analysis change if a_1 and a_2 were replaced with a_{i1} and a_{i2}, respectively, where u_i is independent of $(\mathbf{X}_i, a_{i1}, a_{i2})$?