Advanced Econometrics (PhD in Economics) and Advanced Topics in Econometrics (PhD in Mathematics Applied to Economics and Management)
2025/2026
Exercise Sheet 4 - Limited Dependent Variable Models
(version 11/10/2025)

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

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

where $c > 0$ is a known censoring constant. Note also that

$$\mathbb{E}[\log t_i \mid X_i] = \mathbf{X}_i'\boldsymbol{\beta}_0 - \sigma\phi(a_i) + \left[m - \mathbf{X}_i'\boldsymbol{\beta}_0\right]\left[1 - \Phi(a_i)\right],$$

where $m = \log c$, $a_i = (m - \mathbf{X}_i'\boldsymbol{\beta}_0)$, and $\Phi(a)$ and $\phi(a)$ are, respectively, the cumulative distribution function and density function of the standard normal distribution.

   (a) Consider the following method for estimating $\boldsymbol{\beta}_0$. Using all observations, we run the OLS regression of $\log(t_i)$ on $\mathbf{X}_i$. Explain why this does not generally produce a consistent estimator of $\boldsymbol{\beta}_0$.

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

   (c) 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 $\mathbf{X}_i$).

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

   (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 $Y_i^*|X_i \sim N(\mathbf{X}_i'\boldsymbol{\beta}_0, \sigma^2)$. Note that

$$\mathbb{E}[Y_i \mid \mathbf{X}_i,\, a_1 < Y_i < a_2] = \mathbf{X}_i'\boldsymbol{\beta}_0 + \sigma\,\frac{\phi\left(d_{1i}\right) - \phi\left(d_{2i}\right)}{\Phi\left(d_{2i}\right) - \Phi\left(d_{1i}\right)},$$

where

$$d_{1i} = \frac{a_1 - \mathbf{X}_i'\boldsymbol{\beta}_0}{\sigma}, d_{2i} = \frac{a_2 - \mathbf{X}_i'\boldsymbol{\beta}_0}{\sigma}$$

and $\Phi(a)$ and $\phi(a)$ are, respectively, the cumulative distribution function and density function of the standard normal distribution.

(a) 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 $\boldsymbol{\beta}_0$.

(b) 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.

(c) 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 $\mathbf{X}_i$.

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

(e) 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},$a_{i2})$?