FEN, Universidad de Talca

Discrete Choice Models: Problem Set 3

Professor: Mauricio Sarrias

2020

1 Theory

1. Show that the ML estimator of the Probit model is consistent. (Hint: Use the Theorem of Consistency of conditional ML without compactness from Lecture 1 and the fact that $|\log \Phi(u)| \leq C(1+|u|^2)$, where C is some constant.)

2 Application 1: Mental health and extreme natural events

In this section you will work with the dataset called `data_ee.dta`. This dataset contains variables related to individuals' mental health measured after the 27F earthquake, control variables measured before the earthquake and measures of earthquake/tsunami intensity. Assume you are interested in the "impact" of individuals' mental health due to the 27F earthquake. In particular, the model in mind is the following:

$$h_{i1}^* = \alpha + \beta \ln(y_{i0}) + \gamma r_i + \boldsymbol{x}_{i0}'\boldsymbol{\delta} + \epsilon_{i0}, \quad i = 1, ..., N, \tag{2.1}$$

where h_{i1}^* is a continuous (latent) measure for individual i's mental health measured after the earthquake, so that higher values of h_{i1}^* implies better mental health status; r_i is some measure of the intensity of the earthquake/tsunami that took place on February 27th, 2010; $\ln(y_{i0})$ is the natural logarithm of the monthly household income and $\boldsymbol{x}_{i0}$ is a set of controls measured before the earthquake (2009). However, we do not observed the true mental health, h_{i1}^*, but rather a binary variable, h_{i1}, indicating whether the individual does not have prevalence of post traumatic stress disorder (PTSD) and 0 otherwise. In particular:

$$h_{1i} = \begin{cases} 1 & \text{if } h_{i1}^{*} > 0 \text{ so the individual does not have PTSD} \\ 0 & \text{Otherwise} \end{cases} \tag{2.2}$$

This variable is labeled as `nprev_2010` in the dataset.
Throughout this homework assume that the controls are the following (in stata command); $\boldsymbol{x}_{i0}$:

```
global ind "i.male_2009 age_2009 c.age_2009#c.age_2009 i.hhead_2009 sch_2009  ///
           i.healthp_2009 hsize_2009 i.activ_2009 i.civstat_2009 i.cviv_2009 "
global city "i.coast dens"
```

Thus, the global `ind` is the set of control at the individual level, whereas the global `city` is the set of control at the commune level.
The variables measuring earthquake intensity are: Peak Ground Acceleration (PGA), `pga_w`, the Modified Mercalli Intensity (MMI), `mi_w`, and Peak Ground Velocity (PGV), `pgv_w`. PGA is the highest acceleration in an area, and is measured in rates of gravity acceleration units ($g = 9.8 m/s^2$, so $0.5PGA$ is equivalent to $4.9 m/s^2$), while PGV captures maximum velocity (cm/s). The MMI measure combines both acceleration and velocity with observed (subjective) intensity to generate an estimate of potential damage, which is assigned to a numeric scale. Although both PGA and PGV provide a measure of instrumental intensity, that is, ground shaking recorded by seismic instrument, people are more sensitive to ground acceleration than velocity, whereas structures are more susceptible to velocity than acceleration. MMI quantifies the effects on the Earth's surface, humans, man-made structures, etc, and it is intended to capture how the earthquake was felt by people, and hence, it is a subjective scale of earthquake intensity. All three geological shock variables are calculated at the smallest statistical sampling unit obtainable, and assigned accordingly to each household in the affected area. The following table can be useful for interpretation:

Table 2.1: Ranges of intensity of peak ground acceleration (PGA), velocity (PGV), and Modified Mercalli Intensities (MMI)

Perceived Shaking	Potential Damage	PGA (%g)	PGV (cm/s)	MMI
Not felt	None	<0.0017	<0.1	I
Weak	None	0.0017 – 0.014	0.1 – 1.1	II-–III
Light	None	0.014 – 0.039	1.1 – 3.4	IV
Moderate	Very light	0.039 — 0.092	3.4 – 8.1	V
Strong	Light	0.092 – 0.18	8.1 – 16	VI
Very strong	Moderate	0.18 – 0.34	16 – 31	VII
Severe	Moderate to heavy	0.34 – 0.65	31 – 60	VIII
Violent	Heavy	0.65 – 1.24	60 – 116	IX
Extreme	Very heavy	>1.24	>116	X+

For this homework I highly recommend the following article: Karaca-Mandic, P., Norton, E. C., & Dowd, B. (2012). Interaction terms in nonlinear models. Health services research, 47, 255-274.

1. Estimate model 2.1 using a Probit model with clustered standard errors at the commune level (use `vce(cluster cod_com_casen)`) for each of the three earthquake intensity variable: PGA, PGV and MMI (one model for each earthquake measure). Compute the average marginal effect (and its standard error) for the three measures and interpret the results. Do the AME make sense? How can you compare the AME across the three measures? (Note: to consider the clustered standard errors in the computation of the Delta Method use `vce(unconditional)` in the `margins` command.)

2. Assume that you suspect that the effect of earthquake is on mental health is increasing, so that the model is:

$$h_{i1}^{*} = \alpha + \beta \ln(y_{i0}) + \gamma r_i + \theta r_i^2 + \boldsymbol{x}_{i0}'\boldsymbol{\delta} + \epsilon_{i0}, \quad i = 1, ..., N, \tag{2.3}$$

   Find the average marginal effect of r_i on $\Pr(h_{i1} = 1|\boldsymbol{x}_i)$.

3. Estimate model 2.3 using `mi_w` as your measure of r_i, and compute the AME of r_i for different values of r_i. Plot the AME using command `marginsplot` using the clustered standard errors to compute the 95% CI. What can be interpreted from the results?

4. Suppose you are interested in the compensating variation for being affected by the earthquake. In particular you are interested in *how much additional income an average individual would require to offset the negative impact of the earthquake so she/he can be as well-off as before the shock?* One way of getting such measure is:

$$\overline{MV} = \frac{dy}{dr} = -\frac{\partial h_i/\partial r_i}{\partial h_i/\partial y_i}, \tag{2.4}$$

   where $\overline{MV}$ represents the average marginal value, r_i is earthquake intensity and y_i is household income. Show that the MV using model 2.3 is

$$\overline{MV} = \frac{dy_i}{dr_i} = -\frac{\gamma + 2\theta r_i}{\beta} y_i, \tag{2.5}$$

   How do you interpret 2.5?

5. Compute average $\overline{MV}$ (not MV at the average of y_i) for different values of r_i and plot it with its 95%. Do he results make sense? CI (Hint: use `margins` and `expression` argument. Note that y_i corresponds to `ytd_09` variable in the sample **which is in US dollars**).

6. You suspect that the **marginal effect** of r_i on h_{i1} varies by gender in Equation 2.1. Explain how do you test this hypothesis.

7. Perform your test from previous question using MMI as the earthquake's intensity measure in Equation 2.1 and `margins` command with the AME. What is your conclusion? Why the interaction term is statistically insignificant but the average marginal effects are significant? (Hint: read how to perform a test after using `margins`)

8. Run a Probit model for women and other for men, using Equation 2.1 and MI, and compute the partial effect of r_i. Are those estimates different to the previous question? Why is the difference between this approach and the previous approach in question 7?