Problem Set 2: Discrete Choice

Joan Llull
Conghan Zheng

Microeconometrics
Fall 2024

Introduction

In the following exercises you will estimate the importance of industry-specific human capital on the choice of sector of economic activity after exogenous displacement. For each exercise you should **provide estimates, report standard errors, and briefly interpret the results**. Unless specified differently, you can use pre-programmed STATA or Matlab commands.

Due date October 4th at 12:15 pm, by email to conghan.zheng@uab.cat. You should submit in one zipped folder with:

1. a PDF document containing your analytical solution, if requested in the exercise, any output (figures, tables, etc.) obtained after running the code and your interpretations of the results, instructions on how to run your script(s) if necessary;
2. your code that could run without errors and reproduce the results reported in your explanatory document, specifying the package dependencies if necessary;
3. the name of the zip or rar file, which is `PS2+[your name]`.

Data `PS2.dta` is based on records provided by the *Veneto Workers Histories* (hereafter VWH), a large administrative panel that includes all individuals working in the Italian region of Veneto over the years 1975-2001. VWH provides detailed information on earnings, employment spells, and worker and firm characteristics.

For the reasons explained in the next section, `PS2.dta` contains a subsample of workers who were subject to mass employment reductions in 1999.[1] Mass layoffs are defined with the use of the endogenous separation rate, as in Jacobson, LaLonde and Sullivan (1993) and von Wachter, Song and Manchester (2009).[2] Displaced workers satisfy the requirements imposed by the literature.[3] The final sample contains basic worker characteristics (age, gender, labor

[1] Year 1999 was chosen arbitrarily.

[2] The endogenous separation rate is defined as the ratio of workers who were fired or quit during a quarter over the total amount of workers at the establishment at the beginning of the quarter. A mass layoff (or employment shock) is said to take place when the endogenous separation rate exceeds 0.33 in a firm with more than ten workers.

[3] The sample excludes the agriculture and mining sector. Only employees who were no older than 40 and with tenure longer than four years at the time of separation were taken into account. Since we are interested in permanent separations, the temporary ones were excluded from the sample.

market experience, residence, etc.), duration of unemployment spell, industry prior and after displacement, and industry-specific experience for eight distinctive subgroups: *manufacturing* (light, heavy, electricity and construction), *services* (sales, financial, other) and *public sector* (administration, health). More details about the variables are provided in the Appendix.

Model In this problem set we want to examine the importance of industry-specific human capital for the job choices of displaced workers. Distinguishing exogenous separations will allow us to measure the reliance on industry experience during the distress caused by mass-layoffs. First let's consider a simple binary outcome model where the conditional probability of working after displacement in the same industry as prior to the layoff is

$$p(\boldsymbol{x}) \equiv \Pr[\text{Same industry} = 1|\boldsymbol{x}] = F(\boldsymbol{x}'\boldsymbol{\beta}), \tag{1}$$

where $F(\boldsymbol{x}'\boldsymbol{\beta})$ is a specified function of $\boldsymbol{x}'\boldsymbol{\beta}$ that can either take the logit or probit form. *Same industry* is an indicator that equals 1 if the industry after displacement is the same as before, and 0 otherwise; $\boldsymbol{x}$ contains a set of controls such as labor market experience (`exper`), unemployment duration (`unempl_dur`), age, gender, log wage prior to the layoff (`ln_wage`) and an indicator of current residence (`veneto_resid`). Industries are classified into eight categories as described before. In Equation (1), the variable *experience*, which records the number of years in the labor market, is our proxy for human capital.

Taking advantage of the richness of our data, we can analyze industry-specific human capital and its importance for the choices of displaced workers. The conditional probability of workers choosing alternative j given a vector of characteristics $\boldsymbol{z}$ is

$$p_j(\boldsymbol{z}) \equiv \Pr[\text{Industry} = j|\boldsymbol{z}] = F_j(\boldsymbol{z}'\boldsymbol{\beta}), \quad \text{for } j = 1, ..., m. \tag{2}$$

where j is one of the m alternative industry categories after exogenous displacement and $\sum_{j=1}^{m} p_j = 1$. The vector of regressors $\boldsymbol{z}$ contains both case-invariant variables as in $\boldsymbol{x}$ and alternative-varying variables `e_'industry j'` that record the number of years of experience of worker i in industry j as of displacement. Given (2), we want to estimate the impact of an additional year of experience in industry j on the industry choice of a displaced worker i.

Exercise 1: Binary Outcome Models

In this exercise you will focus on Equation (1):

1. Estimate (1) by specifying a logit and a probit model with the controls in $\boldsymbol{x}$. Interpret the odds ratios for logit.

2. Provide marginal effects for the estimates of the previous point. Interpret the effects of *experience* and *unemployment duration*. Use at-means marginal effects.

Exercise 2: Multinomial Models

In the following exercise you will use Equation (2) to assess the impact of industry-specific human capital on future employment:

1. Consider three industry categories: *manufacturing*, *services* and *public sector*. Using Equation (2) with number of categories $m = 3$, specify a multinomial logit model to estimate the impact of the alternative-invariant variables $\boldsymbol{x}$ on the probability of choosing a particular industry. Interpret the results in terms of relative-risk ratios.

2. Compute at-means marginal effects for the estimates in Exercise 2.1. Interpret them for *experience*, *unemployment duration* and *female*.

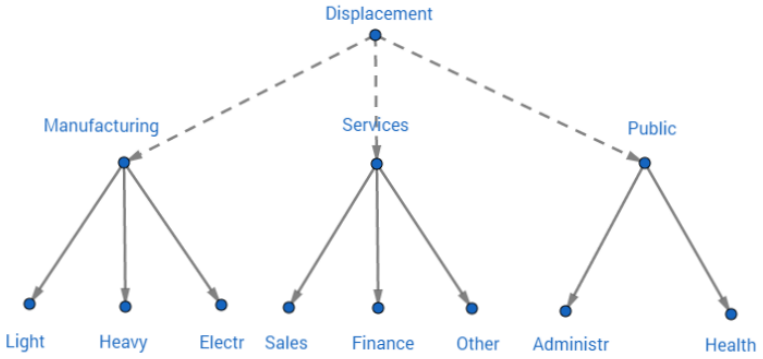

Figure 1: Nested structure of industry choice

Now let's change the number of categories. Consider eight distinctive industries as described above and the nested structure provided by Figure 1.

3. Include the alternative-varying variables $exper_{ij}$ into Equation (2) and use a conditional logit model with eight categories of sub-industries. Interpret the coefficient of the alternative-varying variables, compute their marginal effects, and make a comment on them for a chosen category.

4. Estimate a nested logit model using both alternative-invariant and alternative-varying variables, given the nested structure in Figure 1. Assume that none of the variables influences the first level choice. Compare your results briefly with the ones obtained in Exercise 2.3.

5. Write down the log-likelihood function used in Exercise 2.3 and Exercise 2.4.

References

Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan, "Earnings Losses of Displaced Workers," *The American Economic Review*, 1993, *83* (4), 685–709.

von Wachter, Till, Jae Song, and Joyce Manchester, "Long-Term Earnings Losses due to Mass Layoffs During the 1982 Recession: An Analysis Using U.S. Administrative Data from 1974 to 2004," 2009. Working Paper.

Appendix: Description of variables

- `id`: unique person identifier.
- `year`: year of displacement.
- `age`: age at displacement.
- `female`: gender.
- `ind_prior`: industry category prior to displacement.
- `ind_after`: industry category after displacement.
- `exper`: years of labor market experience.
- `unempl_dur`: unemployment spell after displacement.
- `ln_wage`: log weekly wage in the job prior to displacement (in weeks).
- `veneto_resid`: province of residence (equals 1 if resident of Veneto region and zero otherwise).
- `e_'industry'`: experience in a particular industry category at the time of displacement (in years).