

## Seminar 3. Logit-models

### Task 1. Marketing survey

*File: survey.txt*

Build a logistic model. This data represent the hypothetical results of a marketing survey, where 750 consumers were asked if they would buy a new kind of golf ball based on three pricing points, income, age, family size, gender, and if the consumer had ever used any other product of a certain brand. The results of the survey where coded as a "1" if the respondent answered with a 'yes (=1)' to the purchase decision, otherwise a 'no (=0)'.

Logistic regression was applied to the data and factors such as price, income, and age were found to be statistically significant, along with other variables from the survey. Given this information from an external list provider, the golf retailer could then compute each prospect's propensity to purchase the new golf ball, given their demographic profile and the product's final price offering. A marketing campaign might send special offers to those with the highest probability of purchase.

1. Read your data into EViews, transform it if necessary
2. Examine data for outliers, examine pairwise correlations, examine bivariate correlations & means
3. Select modelling variables
4. Build model and estimate it
5. Examine modelling results
6. Validate model
7. Apply the model to new data
8. Provide an economic analysis of the model (max 2 pages).

### Task 2. Voting decision

*File: school.txt*

Let's study the voting decisions of a sample of individuals in a school budget referendum. The data set contains a sample of 95 individuals. The variables are:

- PUB12 = 1 if 1 or 2 children in public school; = 0 otherwise
- PUB34 = 1 if 3 or 4 children in public school; = 0 otherwise
- PUB5 = 1 if 5 or more children in public school; = 0 otherwise
- PRIV = 1 if 1 or more children in private school; = 0 otherwise
- YEARS = number of years living in the community
- SCHOOL = 1 if individual is employed as a teacher; = 0 otherwise
- LOGINC = logarithm of annual household income (in dollars)
- PTCON = logarithm of property taxes (in dollars) paid per year
- YESVM is a binary variable equal to 1 if the individual voted yes in the election; 0 if individual voted no.

The question of interest is: what factors influence the probability of a yes vote ? This question can be answered by interpreting the estimation results from a logit model.

1. Read your data into EViews, transform it if necessary
2. Examine data for outliers, examine pairwise correlations, examine bivariate correlations & means

3. Select modelling variables
4. Build model and estimate it
5. Examine modelling results
6. Validate model
7. Provide an economic analysis of the model (max 2 pages).

### Task 3. Productivity analysis

File: *productivity.txt*

This problem set concerns predicting productivity of new workers in a large USA manufacturing firm. There are five variables:

- $y_i$  – an observed standardized physical productivity measure for the  $i$ th worker after the initial training period,
- $sex_i$  – a dummy variable for the workers' sex (males are 1),
- $dex_i$  – a score on a physical dexterity exam administered before the worker was hired,
- $lex_i$  – the number of years of education of the worker,
- $quit_i$  – whether the person quit within the first six months (quitters are 1).
- $job\_tenure_i$  - actual duration of employment
- censored – censoring indicator, if the censoring indicator is 0 then the corresponding duration is censored.

1. Estimate the model  $y = \alpha_0 + \alpha_1 sex + \alpha_2 dex + \alpha_3 lex + \alpha_4 lex^2 + u$ .
2. Now consider a similar model for quits

$$\text{logit}(P(\text{quit} = 1)) = (\beta_0 + \beta_1 sex + \beta_2 dex + \beta_3 lex + \beta_4 lex^2)$$

where  $\text{quit} = 1$  if the worker quit within the first 6 months after employment, and is 0 otherwise. Estimate this model by logit, interpret the estimated parameters, in particular the estimated education effect. Change estimation for probit. Compare results.

3. Explore the effect of gender.
4. Presumably, there is a fixed cost of hiring and training so there is an incentive on the part of the firm to avoid hiring workers who are likely to quit after only a few months. How would your findings be expected to influence the firms willingness to hire workers of various education levels?
5. Provide an economic analysis of the model (max 2 pages).

### Task 4. Salary discrimination

File: *salary\_discr.txt*

These are the salary data consisting of observations on six variables for 52 tenure-track professors in a small college. The variables are:

- $sx$  = Sex, coded 1 for female and 0 for male
- $rk$  = Rank, coded
  - 1 for assistant professor,

- 2 for associate professor, and
- 3 for full professor
- yr = Number of years in current rank
- dg = Highest degree, coded 1 if doctorate, 0 if masters
- yd = Number of years since highest degree was earned
- sl = Academic year salary, in dollars.

1. Import data in EViews and adjust non-numerical values.
2. Propose the model to check if the discrimination exists in the college.
3. Estimate the model, provide all necessary tests.
4. Provide an economic analysis of the model.

### Task 5.

*File: uni\_grades.txt*

A researcher is interested in how variables influence the chance to admit the university.

- Admit – is a binary variable, 1 – admitted, 0 – not admitted;
- GRE = Graduate Record Exam scores;
- GPA = grade point average;
- Rank = prestige of the undergraduate institution,

1. Import data in EViews.
2. Propose the model to check if the factors could define the possibility to be admitted.
3. Estimate the model, provide all necessary tests.
4. Provide an economic analysis of the model.

### Task 6.

*File: Agrentina.txt*

The data present the Classification of Counties in Argentina by Urban/Region Categories and Percentage Voting for Peron in 1946 election:

- Region (1=Big City, 2=Township, 3=Countryside)
- Percent Voting for Peron (1=>70, 2=60-70, 3=50-60, 4=40-50, 5=30-40, 6=<30)
- Number of Counties 23-24

1. Import data in EViews.
2. Propose the model and estimate it.