PD Alexander Spermann
Katrin Sommerfeld

Tutorial for Microeconometrics
Albert-Ludwigs-Universität Freiburg
Summer term 2008

Problem Set 4
Treatment Effects (ATT, ATE)

Read in the data set schweiz.raw from the L-drive into TSP. This data set is a subsample of the Swiss Income and Wealth Sample (SEVS = „Schweizerische Einkommens- und Vermögensstichprobe“) and the Socio-medical Indicators for the Population of Switzerland (SOMIPOPS = „Soziomedizinisches Indikatorensystem für die Population der Schweiz“). It contains the following variables in the given order:

emp	Dummy for employment (1 = employed)
NLI	Logarithmized net non-labor income
age	Age
h	Health index (increasing with better health)
educ	Years of education
exper	Potential work experience (Age - education - 7)
LMS	Index for labor market situation, defined as vacancies/ unemployed (on cantonal level)
mar	Dummy for married (1 = married)
kid	Number of children
treat	Treatment-Dummy for tertiary education (1 = more than 13 years of education, i.e. more than highschool)

In the following we will analyze how tertiary education (= treatment) affects women’s employment.

1. Do some descriptive analysis. For example, what are the employment rates of women with and without tertiary education (=treatment)?

2. Under the assumption that treatment and control group differ neither by observable nor by unobservable characteristics, what is the treatment effect?

3. Check whether treatment and control group differ by observable characteristics.

4. Estimate the **"average treatment effect" (ATE).**

Loosely speaking you have to estimate the difference between the employment rates of women with and without tertiary education. More precisely you have to estimate the difference between the two conditional probabilities of working, i.e. $P(emp_i=1|X_i, treat=1) - P(emp_i=1|X_i, treat=0)$ for all individuals.

Procedure:

a) First do a regression of the employment variable on a constant, the potential work experience, the number of children, the logarithmized non-labor income, the labor market situation, health, and the treatment indicator. Write down the estimated equation. Which model do you need to use?

b) Generate two variables for the probability of being employed under the assumption of having or not having some tertiary education. I.e.:

$P(Y_{i,j}|X_i) = \Phi(X_i\beta) = \Phi(cons*\beta_1 + exper_i*\beta_2 + kid_i*\beta_3 + \ldots + j*\beta_7)$ with j=0;1

Hint: To use the cumulative normal distribution: genr varname = cnorm(expression) ;

$$ATE = E(Y_1 - Y_0 \mid X) = E(Y_1 \mid X) - E(Y_0 \mid X)$$

c) Estimate the average difference between $P(Y_{1i}=1|X_i)$ and $P(Y_{0i}=1|X_i)$.

5. Estimate the **"average treatment effect on the treated" (ATT).**

Here, you have to estimate the difference between the two conditional probabilities of working, i.e. $P(Y_{1i}=1|X_i, treat=1) - P(Y_{0i}=1|X_i, treat=1)$, only for treated individuals. Hence, estimate the difference from exercise 4. only for the treated, i.e. those with some tertiary education.

$$ATT = \underbrace{E(Y_1 - Y_0 \mid X, C=1)}_{ATE} = E(Y_1 \mid X, C=1) - E(Y_0 \mid X, C=1)$$

$$\underbrace{\qquad\qquad\qquad\qquad}_{ATT}$$