

Assignment 4

Econometrics I

Universidad Carlos III de Madrid

Gabriel Merlo

```
# install (if missing) packages
list_packages <- c("dplyr", "tidyr")
new_packages <- list_packages[!(list_packages %in% installed.packages()[,"Package"])]
if(length(new_packages)) install.packages(new_packages)

# Load packages
sapply(list_packages, require, character.only = TRUE)
```

Exercise 1

(a) Read the data (`Penn<- as.data.frame(read.table("penn_jae.dat", header=T))`); and estimate the ATE using the standard difference of sample means and a linear regression using as controls.

```
# Load data
penn <- as.data.frame(read.table("penn_jae.dat", header=T))

# Keep control group and treatment group 4
penn4 <- penn %>% filter(tg == 0 | tg == 4)

# Recode treatment variable
penn4$tg <- recode(penn4$tg, `4` = 1L)

# Control variables
x <- "female+black+othrace+dep+q2+q3+q4+q5+q6+age<35+age>54+durable+lusd+husd"

## ATE
#Difference of sample means
diff_mean <- penn4 %>%
  group_by(tg) %>%
  summarize(mean = mean(inuidur1)) %>%
  spread(tg, mean) %>%
  summarize(diff = `1` - `0`)

# Linear regression with controls
ate <- lm(as.formula(paste("inuidur1 ~ tg+", x)), data = penn4)
```

The difference in the mean of log of duration of unemployment between treated and control groups is -0.83. This implies that those that receive the treatment spend less time being unemployed than those who don't get the treatment.

Controlling by observable characteristics of the individuals, the log of duration of unemployment is 0.7 smaller for the individuals that receive the treatment. Once we control by our vector of observables \mathbf{x} , the effect of

the treatment is 0.13 smaller than when comparing using the difference of means.