

## pset 2 Labor

```
#=====#
# ==== Labor Problem set 2 ====
#=====#

#=====#
# ==== load packages and clear data ====
#=====#

# clear data and consol
rm(list = ls(pos = ".GlobalEnv"), pos = ".GlobalEnv")
options(scipen = 999)
cat("\f")

# load packages
library(data.table)
library(lmtest)
library(sandwich)
library(broom)
library(foreign)
library(AER)
library(ivpack)
library(Matrix)
library(knitr)
library(kableExtra)
library(xtable)
#=====#
# ==== data set up ====
#=====#

# load data
dt <- data.table(read.dta("c://users/Nmath_000/Documents/MI_school/Second Year/621 Labor/Assignments/"))

# create needed vars
dt[, age_sq := age^2]
dt[, age_cu := age^3]
dt[, educ_sq := educ^2]
dt[, educ_cu := educ^3]
dt[, age_educ := age*educ]
dt[, age_sq_educ := age_sq*educ]
dt[, age_educ_sq := age*educ_sq]
dt[, nonlab_i := (faminc - wage*hours - huswage*hushrs)/1000]

# create a list to store main results
r_l <- list()

#=====#
# ==== a replicate table 3 ====
#=====#
```

```

# get vars to mean ans sd
var_1 <- c("age", "educ", "kidslt6", "kidsge6", "husage", "huseduc", "wage", "huswage", "nonlab_i", "l

tb1 <- melt.data.table(round(dt[,lapply(.SD,mean), .SDcols =var_1],2), value.name = "Full Sample Mean
tb2 <- melt.data.table(round(dt[,lapply(.SD,sd), .SDcols =var_1],2), value.name = "Full Sample SD")
tb3 <- melt.data.table(round(dt[inlf == 1,lapply(.SD,mean), .SDcols =var_1],2), value.name = "Working
tb4 <- melt.data.table(round(dt[inlf == 1,lapply(.SD,sd), .SDcols =var_1],2), value.name = "Working W

# merge them all
r_1[["a"]] <- Reduce(function(x, y) merge(x, y, by = "variable", all = T), list(tb1, tb2, tb3, tb4))

#####
# ==== b baseline ols ====
#####

# run regression
base_lm <- lm(hours ~ lwage + nwifeinc + kidslt6 + kidsge6 + age + educ , dt)

# get robust standard errors.
lm_robust <- coeftest(base_lm, vcov = vcovHC(base_lm, type="HCO"))

r_1[["b"]] <- data.table(tidy(lm_robust))

#####
# ==== C IV estimate ====
#####

iv_reg <- ivreg(hours ~ lwage + nwifeinc + kidslt6 + kidsge6 + age + educ |
               nwifeinc + kidslt6 + kidsge6 + age + educ +
               age_sq + age_cu + educ_sq + educ_cu +
               age_educ + age_sq_educ + age_educ_sq +
               unem + city + motheduc + fatheduc , data = dt)

# robust the se
iv_reg <- robust.se(iv_reg)

r_1[["c"]] <- data.table(tidy(iv_reg, conf.int = TRUE))

#####
# ==== D repwage ====
#####

# run regression
base_lm <- lm(hours ~ lwage + nwifeinc + kidslt6 + kidsge6 + age + educ , dt[repwage >0,])

# get robust standard errors.
lm_robust <- coeftest(base_lm, vcov = vcovHC(base_lm, type="HCO"))

r_1[["d"]] <- data.table(tidy(lm_robust))

#####

```

```

# ==== E repwage IV ====
#####

iv_reg <- ivreg(hours ~ lwage + nwifeinc + kidslt6 + kidsge6 + age + educ |
               nwifeinc + kidslt6 + kidsge6 + age + educ + repwage , data = dt[repwage >0,])

# robust the se
iv_reg <- robust.se(iv_reg)

r_l[["e"]] <- data.table(tidy(iv_reg, conf.int = TRUE))

#####
# ==== F probit ====
#####

# create variable for working
dt[hours > 0, working := 1]
dt[hours <= 0, working := 0]

# estimate a probit model
myprobit <- glm(working ~nwifeinc + kidslt6 + kidsge6 + age + educ +
               age_sq + age_cu + educ_sq + educ_cu +
               age_educ + age_sq_educ + age_educ_sq +
               unem + city + motheduc + fatheduc,
               family = binomial(link = "probit"),
               data = dt)

r_l[["f"]] <- data.table(tidy(myprobit))

#####
# ==== G H heckman ====
#####

dt[, IMR := dnorm(myprobit$linear.predictors)/pnorm(myprobit$linear.predictors)]

# run regression
samp_sel <- lm( lwage ~ age + educ +
               age_sq + age_cu + educ_sq + educ_cu +
               age_educ + age_sq_educ + age_educ_sq +
               unem + city + motheduc + fatheduc + IMR , dt)

vars <- c("age", "educ", "age_sq", "age_cu", "educ_sq", "educ_cu", "age_educ", "age_sq_educ", "age_educ_sq")

dt[, lapply(.SD, function(y) sum(length(which(is.na(y))))), .SDcols = vars]

# NOTE: the standard errors are wrong since IMR is an estimate. See the worksheet for explanation

r_l[["h"]] <- data.table(tidy(samp_sel))

#####

```

```

# ==== I est hours ====
#####

# get fitted values of lwage for full sample (including non workers ) without the IMR term
B <- as.matrix(r_l[["h"]][term != "IMR", estimate])

X <- dt[, setdiff(r_l[["h"]][, term], c("(Intercept)", "IMR")), with = FALSE]
X[, intercept := 1]
setcolorder(X, c("intercept", setdiff(colnames(X), "intercept")))
X <- as.matrix(X)

fitted <- X%*%B

# add it to data
dt[, lwage_hat := fitted]

# actually I only want estimates for people included in the original regression
dt[is.na(lwage), lwage_hat := NA ]

# estimate hours worked
hrs_reg <- lm( hours ~ lwage_hat + nwifeinc + age + educ + kidslt6 + kidsge6 + IMR , dt)

r_l[["i"]] <- data.table(tidy(hrs_reg))

#####
# ==== J IV and sample correction ====
#####

iv_reg <- ivreg(hours ~ lwage + nwifeinc + kidslt6 + kidsge6 + age + educ + IMR |
  nwifeinc + kidslt6 + kidsge6 + age + educ + IMR +
  age_sq + age_cu + educ_sq + educ_cu +
  age_educ + age_sq_educ + age_educ_sq +
  unem + city + motheduc + fatheduc , data = dt)

# robust the se
iv_reg <- robust.se(iv_reg)

r_l[["j"]] <- data.table(tidy(iv_reg, conf.int = TRUE))

#####
# ==== save tables ====
#####

names(r_l)

save_tex_tables <- function(letter = NULL){

  table <- r_l[[letter]]

  print(xtable(table, type = "latex"),
    file = paste0("C:/Users/Nmath_000/Documents/Code/courses/econ 621/assignment_2_tex/t_", letter,
    include.rownames = FALSE,

```

```
floating = FALSE)
}
lapply(names(r_l), save_tex_tables)
```