section2\_IVA

Eline van Groningen, Paola Priante, Valery Maasdamme, Yuhu Wang

9/25/2020

# Instrumental Variable Analysis: Effect of Compulsory Schooling on Wages

Downloading the libraries

# Load libraries  
library(tidyverse)  
library(stargazer)  
library(dagitty)  
library(gridExtra)  
library(tinytex)  
library(stargazer)  
library(AER)  
library(ivpack)

# Set working director  
setwd("C:/Users/Administrator/Desktop/NewStart/Courses/AdvancedStatisticsandProgramming/assignment2/github/BAM\_ASP\_A2/data")  
  
# Load csv and generate subset containing only variables for interest  
da.IV <- read.csv("IV\_dataset.csv", header = TRUE)  
da.IV <- subset(da.IV, select = c("age", "educ", "lnwage", "married", "qob",   
 "SMSA", "yob"))  
  
## Subset the data set so that we could focus on the variables above according to the order  
da.IV <- read.csv("IV\_dataset.csv", header = TRUE)  
da.IV <- subset(da.IV,select = c("age","educ","lnwage","married","qob","SMSA","yob"))  
## Subset the dataset so that we could focus on the variables above according to the order

stargazer(da.IV,type = "text")  
summary(as.factor(da.IV$married))

# Convert to factor variables  
da.IV$married <- as.factor(da.IV$married)  
da.IV$qob <- as.factor(da.IV$qob)  
da.IV$SMSA <- as.factor(da.IV$SMSA)  
da.IV$yob <- as.factor(da.IV$yob)  
  
# To change those variables which should be factor variables into factor variables  
g1.1 <- ggplot(data = da.IV, aes(qob, educ)) +  
 geom\_point(size = 0.5) +   
 geom\_smooth(method = "lm", color = "blue", alpha = 0.2) +   
 theme\_bw() +  
 labs(caption = "Figure 2.1") +  
 geom\_boxplot() +   
 theme(plot.caption = element\_text(hjust = 0.5, size = 12, face = "bold")) +   
 labs(x = "Quarter of Birth", y = "Education(in years)")  
g1.1

rsltIV <- ivreg(lnwage ~ educ|qob,data = da.IV)  
summary(rsltIV, diagnostics = TRUE)

library(ivreg)  
rslt2SLS.A <- ivreg(lnwage ~ educ | qob, data=da.IV)  
summary(rslt2SLS.A)  
stargazer(rslt2SLS.A, type= "text")

rslt2SLS.B <- ivreg(lnwage ~ educ + married + SMSA | married + SMSA + qob,   
 data=da.IV)  
summary(rslt2SLS.A)  
stargazer(rslt2SLS.A, rslt2SLS.B)

#Robust standard errors  
modelIV <- ivreg(lnwage ~ educ + married + SMSA | married + SMSA + qob ,   
 data=da.IV)  
summary(modelIV)  
  
#Standard errors (superfluous in the case of seBasic)  
seBasic <- sqrt(diag(vcov(modelIV)))  
seWhite <- sqrt(diag(vcovHC(modelIV , type="HC0")))  
library(vcov)  
# Make table with stargazer  
stargazer(modelIV , modelIV ,align=TRUE , no.space=TRUE ,intercept.bottom = FALSE ,se = list(seBasic , seWhite), type= "text")

da.IV\_sub <- subset(da.IV,select = c("age", "educ", "lnwage", "married", "qob",   
 "SMSA", "yob"))  
  
# Convert to factor variables  
da.IV\_sub$married <- as.factor(da.IV\_sub$married)  
da.IV\_sub$qob <- as.factor(da.IV\_sub$qob)  
da.IV\_sub$SMSA <- as.factor(da.IV\_sub$SMSA)  
da.IV\_sub$yob <- as.factor(da.IV\_sub$yob)  
  
# Define OLS models  
rsltOLS.A <- lm(lnwage ~ educ, data=da.IV\_sub)  
rsltOLS.B <- lm(lnwage ~ educ + married + SMSA, data=da.IV\_sub)  
  
# Define IV model   
rsltSLS.A <- ivreg(lnwage ~ educ | qob, data=da.IV\_sub)  
rsltSLS.B <- ivreg(lnwage ~ educ + married + SMSA | married + SMSA + qob,   
 data=da.IV\_sub)  
rsltSLS.C <- ivreg(lnwage ~ educ + married + SMSA | married + SMSA + age + qob,   
 data=da.IV\_sub)  
  
# Generate table containing both models  
stargazer(rsltOLS.A, rsltOLS.B, rsltSLS.A, rsltSLS.B, rsltSLS.C, type="text")  
  
# Test for violation over-identification  
summary(rsltSLS.A, diagnostics = TRUE)  
summary(rsltSLS.B, diagnostics = TRUE)  
summary(rsltSLS.C, diagnostics = TRUE)