Applied Micro PSET 1

Matthew Zhao

2023-01-11

```
library(tidyverse)
library(haven)
library(stargazer)
library(xtable)
options(scipen=999)
```

Question 1

a)

```
df <- read_stata('data/CARD.DTA')
```

$$lwage_i = \beta_0 + \beta_1 educ_i + \beta_3 KWW_i + U_i (1)$$

$$lwage_i = \beta_0 + \beta_1 HS_i + \beta_2 College_i + \beta_3 KWW_i + U_i$$

```
df <- df %>%
  mutate(HS = ifelse((educ >= 12 & educ < 16),1,0),
         College = ifelse(educ >= 16,1,0))

m1 <- lm(lwage ~ educ + KWW, data = df)
m2 <- lm(lwage ~ HS + College + KWW, data = df)
models <- list(m1,m2)
stargazer(models,type='text',digits=3,
          title='OLS Estimates for Equations (1) and (2)')
```

```
##
## OLS Estimates for Equations (1) and (2)
## =======================================================================
##                                   Dependent variable:
##                     ---------------------------------------------------
##                                          lwage
##                              (1)                         (2)
## ---------------------------------------------------------------------
## educ                      0.021***
##                           (0.003)
##
```

```
## HS                                                        0.125***
##                                                           (0.022)
##
## College                                                    0.195***
##                                                           (0.026)
##
## KWW                              0.019***                  0.019***
##                                 (0.001)                   (0.001)
##
## Constant                         5.351***                  5.514***
##                                 (0.039)                   (0.031)
##
## -------------------------------------------------------------------
## Observations                      2,963                     2,963
## R2                                0.196                     0.199
## Adjusted R2                       0.195                     0.199
## Residual Std. Error        0.396 (df = 2960)          0.395 (df = 2959)
## F Statistic          360.893*** (df = 2; 2960) 245.547*** (df = 3; 2959)
## ===================================================================
## Note:                                       *p<0.1; **p<0.05; ***p<0.01
```

For (1), the interpretation of β_1 is that holding KWW (a proxy for intelligence) fixed, we expect an average approximate increase of 2.1% in wages for each additional year of education. We interpret β_3 as an expected approximate increase of 1.9% in wages for each additional increase in KWW score holding years of education fixed.

For (2), the interpretation of β_1 is that holding KWW (a proxy for intelligence) fixed, we expect an average approximate increase of 12.5% in wages for an individual with only a high school degree. We interpret β_2 as an expected approximate increase of 19.5% in wages for an individual with a college degree compared to those with no degree holding KWW score fixed. We interpret β_3 as an expected approximate increase of 1.9% in wages for each additional increase in KWW score holding education level/degree status fixed.

b)

The OLS estimator $\hat{\beta}_1$ from (1) is biased because it is unlikely that the assumption $E[U|educ] = 0$ since individuals may select into differing amounts of education based on unobservables such as cultural values or family wealth.

c)

Ideally we would want to measure intelligence (KWW) prior to having any schooling. Otherwise, if we measured intelligence after individuals received schooling, it is likely that MLR.4 will be violated since education could have influenced intelligence.

d)

```
m1 <- lm(lwage ~ educ + HS + College + KWW, data = df)
m2 <- lm(lwage ~ educ*HS + educ*College + KWW, data = df)

models <- list(m1,m2)
stargazer(models,type='text',digits=3,
          title='OLS Estimates for Additional Specifications')
```

```
## 
## OLS Estimates for Additional Specifications
## =============================================================================
##                                  Dependent variable:
##                      -----------------------------------------------------
##                                         lwage
##                              (1)                          (2)
## ---------------------------------------------------------------------------
## educ                        0.006                        0.012
##                            (0.007)                      (0.011)
## 
## HS                        0.107***                      0.395**
##                            (0.030)                      (0.157)
## 
## College                   0.155***                      -0.461
##                            (0.052)                      (0.291)
## 
## KWW                       0.018***                      0.018***
##                            (0.001)                      (0.001)
## 
## educ:HS                                                 -0.024*
##                                                         (0.014)
## 
## educ:College                                            0.034*
##                                                         (0.019)
## 
## Constant                  5.464***                      5.408***
##                            (0.065)                      (0.102)
## 
## ---------------------------------------------------------------------------
## Observations                2,963                        2,963
## R2                          0.200                        0.202
## Adjusted R2                 0.198                        0.201
## Residual Std. Error    0.395 (df = 2958)            0.395 (df = 2956)
## F Statistic          184.337*** (df = 4; 2958) 124.921*** (df = 6; 2956)
## ===========================================================================
## Note:                                    *p<0.1; **p<0.05; ***p<0.01
```

I prefer model specification (2). This is because when adding education to this model, we see that the coefficient is insignificantly different from zero. This indicates that the degree premia account for the majority of gains to wages from education holding intelligence fixed. **get rid of m2?**

e)

```
m1 <- lm(lwage ~ educ + KWW, data = df %>% filter(age == 24))
m2 <- lm(lwage ~ educ + KWW, data = df %>% filter(age == 28))
m3 <- lm(lwage ~ educ + KWW, data = df %>% filter(age == 32))

models <- list(m1,m2,m3)
stargazer(models,type='text',digits=3,
          title='OLS Estimates for (1) at fixed ages')
```

```
## 
## OLS Estimates for (1) at fixed ages
## ===========================================================================
##                                     Dependent variable:
##                    --------------------------------------------------------
##                                            lwage
##                        (1)                  (2)                  (3)
## ---------------------------------------------------------------------------
## educ                 -0.0005             0.036***             0.043***
##                      (0.010)             (0.010)              (0.011)
## 
## KWW                  0.010***            0.011***             0.015***
##                      (0.003)             (0.003)              (0.004)
## 
## Constant             5.791***            5.455***             5.319***
##                      (0.122)             (0.137)              (0.134)
## 
## ---------------------------------------------------------------------------
## Observations           388                 309                  210
## R2                    0.037               0.122                0.269
## Adjusted R2           0.032               0.116                0.262
## Residual Std. Error  0.400 (df = 385)    0.380 (df = 306)     0.380 (df = 207)
## F Statistic       7.372*** (df = 2; 385) 21.236*** (df = 2; 306) 38.071*** (df = 2; 207)
## ===========================================================================
## Note:                                         *p<0.1; **p<0.05; ***p<0.01
```

```r
m1 <- lm(lwage ~ HS + College + KWW, data = df %>% filter(age == 24))
m2 <- lm(lwage ~ HS + College + KWW, data = df %>% filter(age == 28))
m3 <- lm(lwage ~ HS + College + KWW, data = df %>% filter(age == 32))

models <- list(m1,m2,m3)
stargazer(models,type='text',digits=3,
          title='OLS Estimates for (2) at fixed ages')
```

```
## 
## OLS Estimates for (2) at fixed ages
## ===========================================================================
##                                     Dependent variable:
##                    --------------------------------------------------------
##                                            lwage
##                        (1)                  (2)                  (3)
## ---------------------------------------------------------------------------
## HS                    0.031               0.248***             0.412***
##                      (0.063)             (0.082)              (0.080)
## 
## College              -0.007               0.389***             0.479***
##                      (0.076)             (0.090)              (0.095)
## 
## KWW                  0.010***            0.010***             0.013***
##                      (0.003)             (0.003)              (0.004)
## 
## Constant             5.766***            5.702***             5.608***
##                      (0.083)             (0.116)              (0.123)
## 
```

```
## ---------------------------------------------------------------------------------
## Observations              388                  309                  210
## R2                        0.039                0.142                0.313
## Adjusted R2               0.031                0.133                0.303
## Residual Std. Error   0.400 (df = 384)     0.376 (df = 305)     0.369 (df = 206)
## F Statistic          5.148*** (df = 3; 384) 16.765*** (df = 3; 305) 31.232*** (df = 3; 206)
## =================================================================================
## Note:                                              *p<0.1; **p<0.05; ***p<0.01
```

These estimates are generally larger than the estimates from part (a) for ages 28 and 32 and smaller for age 24. They differ because we're conditioning on age **something about OVB and U**.

f)

```
m1 <- lm(lwage ~ educ + KWW, data = df %>% filter(exper == 4))
m2 <- lm(lwage ~ educ + KWW, data = df %>% filter(exper == 8))
m3 <- lm(lwage ~ educ + KWW, data = df %>% filter(exper == 12))

models <- list(m1,m2,m3)
stargazer(models,type='text',digits=3,
        title='OLS Estimates for (1) at fixed experience')
```

```
##
## OLS Estimates for (1) at fixed experience
## ===============================================================================
##                                      Dependent variable:
##                         -------------------------------------------------------
##                                              lwage
##                              (1)                (2)                (3)
## -------------------------------------------------------------------------------
## educ                      0.074***           0.052***           0.108***
##                           (0.026)            (0.012)            (0.019)
##
## KWW                       0.018***           0.014***           0.009*
##                           (0.005)            (0.003)            (0.005)
##
## Constant                  4.498***           5.063***           4.739***
##                           (0.369)            (0.121)            (0.162)
##
## -------------------------------------------------------------------------------
## Observations                163                308                132
## R2                         0.208              0.283              0.452
## Adjusted R2                0.198              0.278              0.444
## Residual Std. Error   0.384 (df = 160)   0.351 (df = 305)   0.366 (df = 129)
## F Statistic          21.024*** (df = 2; 160) 60.154*** (df = 2; 305) 53.227*** (df = 2; 129)
## ===============================================================================
## Note:                                              *p<0.1; **p<0.05; ***p<0.01
```

```
m1 <- lm(lwage ~ HS + College + KWW, data = df %>% filter(exper == 4))
m2 <- lm(lwage ~ HS + College + KWW, data = df %>% filter(exper == 8))
m3 <- lm(lwage ~ HS + College + KWW, data = df %>% filter(exper == 12))
```

```
models <- list(m1,m2,m3)
stargazer(models,type='text',digits=3,
          title='OLS Estimates for (2) at fixed experience')
```

```
##
## OLS Estimates for (2) at fixed experience
## =========================================================================
##                                      Dependent variable:
##                 --------------------------------------------------------
##                                           lwage
##                        (1)                 (2)                 (3)
## -------------------------------------------------------------------------
## HS                   -0.117*             0.160**             0.319***
##                      (0.069)             (0.067)             (0.088)
##
## College                                  0.362***            0.592***
##                                          (0.088)             (0.136)
##
## KWW                  0.021***            0.015***            0.016***
##                      (0.005)             (0.003)             (0.004)
##
## Constant             5.612***            5.540***            5.551***
##                      (0.171)             (0.092)             (0.125)
##
## -------------------------------------------------------------------------
## Observations           163                 308                 132
## R2                    0.183               0.278               0.412
## Adjusted R2           0.172               0.271               0.398
## Residual Std. Error 0.391 (df = 160)    0.353 (df = 304)    0.381 (df = 128)
## F Statistic    17.861*** (df = 2; 160) 39.020*** (df = 3; 304) 29.861*** (df = 3; 128)
## =========================================================================
## Note:                                        *p<0.1; **p<0.05; ***p<0.01
```