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')
```

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

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

```
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 intelligence would be affected by education. In the data, the exam is administered in 1966, so intelligence was measured earlier in life for some individuals and later in life for others.

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.

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
```

```
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. This makes sense since generally experience and education increase with age, thus likely increasing log wages. The estimates differ from those in part (a) because the specification did not include age or any proxy for age, meaning that our estimates were likely biased. Additionally, the increasing $\hat{\beta}_1$ indicates that there are potentially increasing returns to education with respect to age so not having interaction terms also can explain why the estimates are different. Lastly, since many of the observations are in the lowest age group, with lower age being associated with a smaller estimated β_1, this distribution in age in the sample would also bias the coefficient from part (a) downwards.

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
```

Similarly to (e), the estimates increase with experience. This is likely because greater experience increases returns to education, or more specifically, at higher levels of experience, education makes a greater impact on the rate of change in log wage. The estimates differ from those in part (a) since the model did not account for experience and potentially heterogeneous returns to education.

g)

$$lwage_i = \beta_0 + \beta_1 educ_i + \beta_3 KWW_i + \beta_4 exper_i + U_i \quad (1)$$

$$lwage_i = \beta_0 + \beta_1 HS_i + \beta_2 College_i + \beta_3 KWW_i + \beta_4 exper_i + U_i$$

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

```
##
## OLS Estimates for Equations (1) and (2) with experience
## =======================================================================
##                                     Dependent variable:
##                    ----------------------------------------------------
##                                           lwage
##                             (1)                        (2)
## ----------------------------------------------------------------------
## educ                     0.056***
##                          (0.004)
##
## HS                                                  0.237***
##                                                     (0.024)
##
## College                                             0.399***
##                                                     (0.033)
##
## KWW                      0.014***                   0.015***
##                          (0.001)                    (0.001)
##
## exper                    0.027***                   0.022***
##                          (0.003)                    (0.002)
##
## Constant                 4.820***                   5.309***
##                          (0.063)                    (0.037)
##
## ----------------------------------------------------------------------
## Observations              2,963                      2,963
## R2                        0.225                      0.223
## Adjusted R2               0.225                      0.222
## Residual Std. Error   0.389 (df = 2959)         0.389 (df = 2958)
## F Statistic        286.844*** (df = 3; 2959) 212.510*** (df = 4; 2958)
## =======================================================================
## Note:                                     *p<0.1; **p<0.05; ***p<0.01
```

We are testing the null that $\beta_4 = 0$ against the alternative that it is not equal to zero.

After adding *exper* to the models we see that the coefficients become larger (more positive). If these specifications allow us to interpret $\hat{\beta}_1$ causally, we can say that adding *exper* would make us conclude that the causal effect of education on wages is larger than we previously thought and that our previous estimates of (1) and (2) suffered from omitted variable bias causing them to be too small.

h)

$$lwage_i = \beta_0 + \beta_1 educ_i + \beta_3 KWW_i + \beta_4 exper_i + \beta_5 exper_i * educ + U_i \quad (1)$$

$$lwage_i = \beta_0 + \beta_1 HS_i + \beta_2 College_i + \beta_3 KWW_i + \beta_4 exper_i + \beta_6 exper_i * HS + \beta_7 exper_i * College + U_i \quad (2)$$

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

```
##
## Equations (1) and (2) with experience interactions
## ================================================================================
##                                       Dependent variable:
##                     ----------------------------------------------------
##                                              lwage
##                               (1)                         (2)
## --------------------------------------------------------------------------------
## educ                        0.031***
##                             (0.007)
##
## HS                                                      0.037
##                                                        (0.069)
##
## College                                                 0.043
##                                                        (0.070)
##
## KWW                         0.013***                  0.015***
##                             (0.001)                    (0.001)
##
## exper                       -0.008                      0.005
##                             (0.007)                    (0.005)
##
## educ:exper                  0.003***
##                             (0.001)
##
## HS:exper                                               0.015***
##                                                        (0.005)
##
## College:exper                                          0.042***
##                                                        (0.007)
##
## Constant                    5.164***                  5.549***
##                             (0.092)                    (0.066)
##
## --------------------------------------------------------------------------------
## Observations                 2,963                      2,963
## R2                           0.232                      0.234
## Adjusted R2                  0.231                      0.233
## Residual Std. Error    0.387 (df = 2958)          0.387 (df = 2956)
## F Statistic        223.396*** (df = 4; 2958) 150.753*** (df = 6; 2956)
## ================================================================================
## Note:                                        *p<0.1; **p<0.05; ***p<0.01
```

Here we are testing the null that $\beta_5 = 0$ for (1) and $\beta_6 = 0$ and $\beta_7 = 0$ for (2) against the null that they are not zero.

We do see evidence that the wage return to education differs by potential experience with the coefficients rejecting the null at the 1% significance level for both models.

i)

```r
q1i <- df %>%
  mutate(educat = ifelse(educ >= 12,
                         ifelse(educ >= 16,'College','HS'),
                         'No Degree')) %>%
  group_by(educat,exper) %>%
  summarise(mean_lwage = mean(lwage),
            mean_kww = mean(KWW),
            mean_exper = mean(exper)) %>%
  mutate(modela =
           0.125*ifelse(educat == 'HS',1,0) +
           0.195*ifelse(educat == 'College',1,0) +
           0.019*mean_kww +
           5.514,
         modelg =
           0.237*ifelse(educat == 'HS',1,0) +
           0.399*ifelse(educat == 'College',1,0) +
           0.015*mean_kww +
           0.022*mean_exper +
           5.309,
         modelh =
           0.037*ifelse(educat == 'HS',1,0) +
           0.043*ifelse(educat == 'College',1,0) +
           0.015*mean_kww +
           0.005*mean_exper +
           0.015*ifelse(educat == 'HS',1,0)*mean_exper +
           0.042*ifelse(educat == 'College',1,0)*mean_exper +
           5.549) %>%
  pivot_longer(cols = c(mean_lwage,modela,modelg,modelh),
               names_to = 'yvals',
               values_to='values',
               values_drop_na = T)

ggplot(data = q1i, mapping = aes(x = mean_exper, y = values, color = yvals)) +
  geom_point() +
  geom_line() +
  facet_wrap(~educat) +
  theme_light() +
  ggtitle("Model Fit across Levels of Education") +
  labs(x = 'Experience', y = 'Log Wage') +
  theme(legend.title= element_blank())
```

Model Fit across Levels of Education
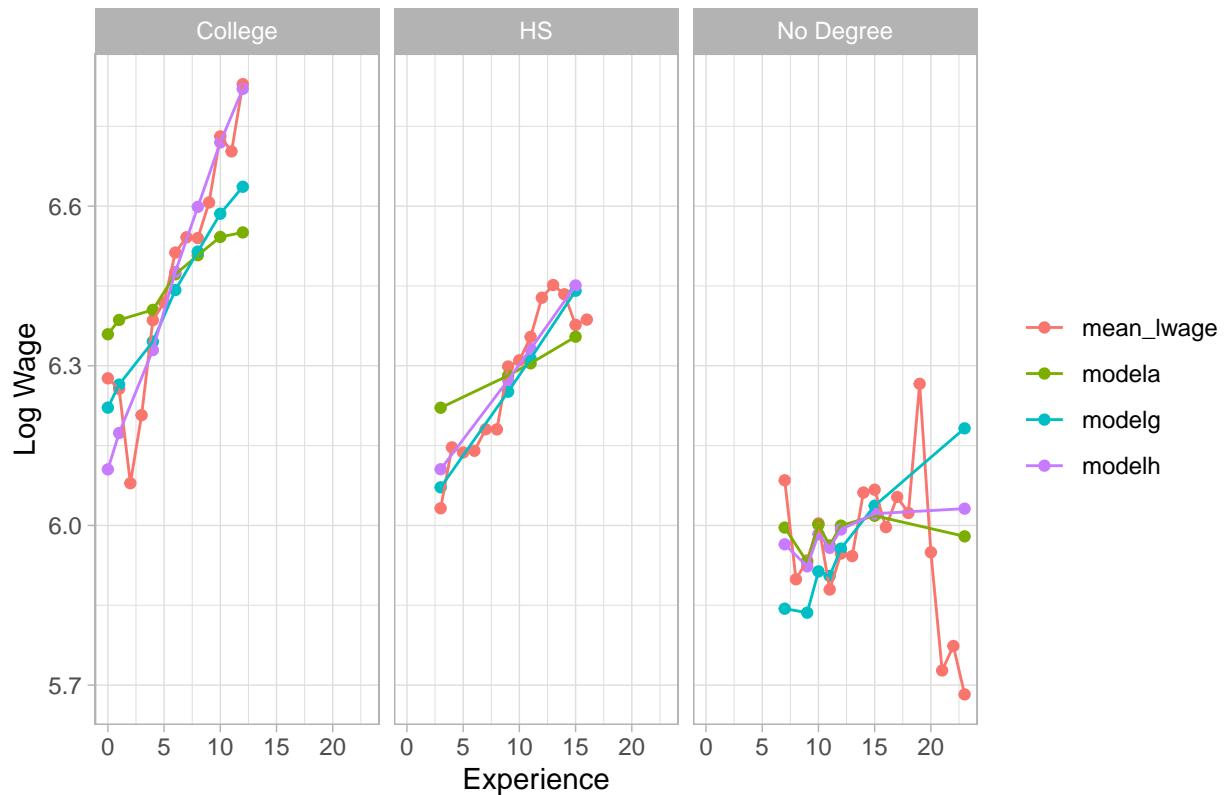
We see that as the model becomes more complex, we are better able to capture variation in wage-experience profiles across the different levels of education. As a result, it appears that the most complex model with experience interactions, model h, best fits the data across the three levels.

j)

(i)

$$lwage_i = \beta_0 + \beta_1 educ_i + \beta_3 KWW_i + \beta_4 black_i + U_i \quad (1)$$

$$lwage_i = \beta_0 + \beta_1 HS_i + \beta_2 College_i + \beta_3 KWW_i + \beta_4 black_i + U_i \quad (2)$$

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

models <- list(m1,m2)
stargazer(models,type='text',digits=3,
          title='Population Models (1) and (2) with race')
```

(ii)

##

```
## Population Models (1) and (2) with race
## ================================================================
##                            Dependent variable:
##                  -------------------------------------------------
##                                      lwage
##                        (1)                        (2)
## ----------------------------------------------------------------
## educ                  0.020***
##                       (0.003)
##
## HS                                              0.113***
##                                                 (0.021)
##
## College                                         0.180***
##                                                 (0.025)
##
## KWW                   0.016***                  0.016***
##                       (0.001)                   (0.001)
##
## black                -0.133***                 -0.129***
##                       (0.019)                   (0.019)
##
## Constant              5.490***                  5.640***
##                       (0.044)                   (0.036)
##
## ----------------------------------------------------------------
## Observations            2,963                     2,963
## R2                      0.209                     0.211
## Adjusted R2             0.208                     0.210
## Residual Std. Error  0.393 (df = 2959)        0.392 (df = 2958)
## F Statistic      260.601*** (df = 3; 2959) 198.149*** (df = 4; 2958)
## ================================================================
## 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% in wages for each additional year of education. We interpret β_3 as an expected approximate increase of 1.6% in wages for each additional increase in KWW score holding years of education fixed. We interpret β_4 as an expected approximate decrease of 13.3% in wages for an individual who is black compared with one who is not holding years of education and intelligence fixed.

For (2), the interpretation of β_1 is that holding KWW (a proxy for intelligence) fixed, we expect an average approximate increase of 11.3% in wages for an individual with only a high school degree. We interpret β_2 as an expected approximate increase of 18% 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 1.6% in wages for each additional increase in KWW score holding education level/degree status fixed. We interpret β_4 as an expected approximate decrease of 12.9% in wages for an individual who is black compared with one who is not holding education level/degree status and intelligence fixed.

(iii) Here we are testing the null hypothesis stating that the coefficient for *black* is 0 ($\beta_4 = 0$) against the alternative that it is nonzero.

Given that the coefficients for both (1) and (2) are large and significantly different from zero, we can conclude that wages do differ by race. However, we cannot say that this is necessarily a causal relationship i.e. that race has a causal impact on wages since it could be the case that other determinants of wage that are not observed here e.g. family wealth are related to race, resulting in selection.

k)

(i)

$$lwage_i = \beta_0 + \beta_1 educ_i + \beta_3 KWW_i + \beta_4 black_i + \beta_5 black_i * educ_i + U_i (1)$$

$$lwage_i = \beta_0 + \beta_1 HS_i + \beta_2 College_i + \beta_3 KWW_i + \beta_4 black_i + \beta_5 black_i * HS + \beta_6 black_i * College + U_i (2)$$

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

models <- list(m1,m2)
stargazer(models,type='text',digits=3,
          title='OLS Estimates including race interaction')
```

(ii)

```
##
## OLS Estimates including race interaction
## ==============================================================================
##                                      Dependent variable:
##                          -----------------------------------------------------
##                                             lwage
##                                  (1)                        (2)
## ------------------------------------------------------------------------------
## educ                           0.016***
##                                (0.003)
##
## HS                                                        0.096***
##                                                           (0.027)
##
## College                                                  0.148***
##                                                           (0.030)
##
## KWW                            0.016***                   0.016***
##                                (0.001)                    (0.001)
##
## black                         -0.366***                  -0.173***
##                                (0.085)                    (0.037)
##
## educ:black                     0.019***
##                                (0.007)
##
## HS:black                                                    0.033
##                                                           (0.042)
##
## College:black                                            0.162***
##                                                           (0.058)
##
## Constant                       5.551***                   5.661***
##                                (0.049)                    (0.039)
##
```

```
## ----------------------------------------------------------------------
## Observations                 2,963                         2,963
## R2                           0.211                         0.214
## Adjusted R2                  0.210                         0.212
## Residual Std. Error    0.392 (df = 2958)          0.392 (df = 2956)
## F Statistic         197.861*** (df = 4; 2958) 133.772*** (df = 6; 2956)
## ======================================================================
## 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) and race fixed, we expect an average approximate increase of 1.6% in wages for each additional year of education. We interpret β_3 as an expected approximate increase of 1.6% in wages for each additional increase in KWW score holding years of education and race fixed. We interpret β_4 as the expected approximate decrease of 36.6% in wages for an individual who is black compared with one who is not holding years of education and intelligence fixed. We interpret β_5 as an expected approximate additional increase of 1.9% in wages for each additional year of education for an individual who is black compared with one who is not holding years of education and intelligence fixed.

For (2), the interpretation of β_1 is that holding KWW (a proxy for intelligence) fixed, we expect an average approximate increase of 9.6% in wages for an individual with only a high school degree. We interpret β_2 as an expected approximate increase of 14.8% 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 1.6% in wages for each additional increase in KWW score holding education level/degree status fixed. We interpret β_4 as an expected approximate decrease of 17.3% in wages for an individual who is black compared with one who is not holding education level/degree status and intelligence fixed. We interpret β_5 as an expected approximate additional increase of 3.3% in wages for an individual with only a high school degree if that individual is black compared with one who is not holding years of education and intelligence fixed. We interpret β_6 as an expected approximate additional increase of 16.2% in wages for an individual with a college degree if that individual is black compared with one who is not holding years of education and intelligence fixed.

(iii) We are testing the null hypothesis that $\beta_6 = 0$ for (1) and $\beta_6 = 0$ and $\beta_7 = 0$ for (2) against the alternative that the coefficients are not zero.

We see evidence that wage return to education differs by race in (1) where the interaction term between race and education is significantly different from zero. We may expect this since black individuals are typically disadvantaged in the workforce, with education playing a role in leveling the playing field and evening out wage disparities.

There is also some evidence in specification (2) that wage return to education differs by race since the estimate of β_7 is large and significantly different from zero while the estimate of β_6 is small and insignificantly different from zero. We may expect this difference because obtaining a college degree opens many opportunities in the form of DEI and various other workplace initiatives as well as careers that would potentially allow individuals who are black to break out of a poverty cycle if we believe individual level poverty traps exist.

l)

In this setting I would prefer the specification from part (k) since it captures significant effects via interaction terms that model j fails to account for, specifically the differential impact of college on black individuals in this sample. Additionally, there are significant differences in the estimated coefficients across these specifications, indicating potential misspecification and OVB, especially in model a.

13

m)

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


models <- list(m1,m2)
stargazer(models,type='text',digits=3,
          title='Sensitivity Test to Location - south')
```

```
## 
## Sensitivity Test to Location - south
## =========================================================================
##                                   Dependent variable:
##                     -----------------------------------------------------
##                                           lwage
##                              (1)                         (2)
## -------------------------------------------------------------------------
## educ                      0.014***
##                            (0.003)
## 
## HS                                                     0.075***
##                                                        (0.027)
## 
## College                                               0.129***
##                                                        (0.030)
## 
## KWW                       0.015***                    0.015***
##                            (0.001)                     (0.001)
## 
## black                     -0.304***                   -0.130***
##                            (0.085)                     (0.036)
## 
## south                     -0.140***                   -0.140***
##                            (0.016)                     (0.016)
## 
## educ:black                0.017***
##                            (0.007)
## 
## HS:black                                                0.036
##                                                        (0.042)
## 
## College:black                                         0.161***
##                                                        (0.057)
## 
## Constant                  5.644***                    5.748***
##                            (0.049)                     (0.039)
## 
## -------------------------------------------------------------------------
## Observations                2,963                       2,963
## R2                          0.232                       0.234
## Adjusted R2                 0.231                       0.233
## Residual Std. Error   0.387 (df = 2957)           0.387 (df = 2955)
```

```
## F Statistic           178.616*** (df = 5; 2957) 129.272*** (df = 7; 2955)
## ===========================================================================
## Note:                                        *p<0.1; **p<0.05; ***p<0.01
```

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

models <- list(m1,m2)
stargazer(models,type='text',digits=3,
          title='Sensitivity Test to Location - smsa')
```

```
##
## Sensitivity Test to Location - smsa
## ===========================================================================
##                                   Dependent variable:
##                        ----------------------------------------------------
##                                         lwage
##                             (1)                      (2)
## ---------------------------------------------------------------------------
## educ                     0.013***
##                          (0.003)
##
## HS                                                0.082***
##                                                   (0.027)
##
## College                                           0.130***
##                                                   (0.030)
##
## KWW                      0.015***                 0.015***
##                          (0.001)                  (0.001)
##
## black                   -0.319***                -0.168***
##                          (0.084)                  (0.036)
##
## smsa                     0.155***                 0.157***
##                          (0.016)                  (0.016)
##
## educ:black               0.014**
##                          (0.007)
##
## HS:black                                            0.014
##                                                   (0.042)
##
## College:black                                     0.139**
##                                                   (0.057)
##
## Constant                 5.511***                 5.602***
##                          (0.048)                  (0.038)
##
## ---------------------------------------------------------------------------
## Observations              2,963                    2,963
## R2                        0.235                    0.238
## Adjusted R2               0.234                    0.237
## Residual Std. Error  0.386 (df = 2957)        0.386 (df = 2955)
```

```
## F Statistic           181.811*** (df = 5; 2957) 132.090*** (df = 7; 2955)
## =======================================================================
## Note:                                             *p<0.1; **p<0.05; ***p<0.01
```

It appears that our conclusions in part l are robust to specifications including location controls e.g. south and SMSA, with the coefficients seeing little change when including these controls.

Question 2

a)

While MLR.1-MLR.4 hold for the population model (3), MLR.4 may not hold for (4). This is because while $E[U|X_1] = 0$ holds in (3) by assumption, we cannot be certain this is the case in (4) i.e. that MLR.4 holds where $E[\hat{U}|X_2] = 0$ because $\hat{U}$ contains U and X_2, and X_1 and X_2 are likely to be related to each other. As a result, leaving out X_2 means that MLR.4 is likely violated and our estimate of β_1 is biased since (4) suffers from OVB. Thus, we cannot say that $\hat{\beta}_1$ is an unbiased and consistent estimate of the causal effect of X_1 on Y.

b)

As mentioned in part (a), MLR.4 may not plausible in (4) since X_2 is omitted from the model. In order for $\hat{\beta}_1$ to be an unbiased and consistent estimate of β_1, we would need to assume that X_1 and X_2 are uncorrelated. This is most likely not credible since people typically choose partners who come from similar socioeconomic backgrounds and with similar income.

c)

We would need data on X_2, that is, we need data about the incomes of these head of households' partners. Specifically, we would want to obtain data that shows that their incomes are not correlated or simply change the specification to include X_2 and obtain the new OLS estimates with the new data to better reflect the population model.

Question 3

A paper that I am interested in replicating is this paper titled "Employment Effects of Financial Constraints during the Great Recession" by Michael Siemer which explores how financial constraints on firms after the Great Recession led to higher unemployment in the recovery period. This paper interests me since it illustrates a macrofinance explanation for the unusually slow labor market recovery post-2008 using firm-level data. The author employs a differences-in-differences identification strategy which I anticipate will be covered later in the course and replication data can be found **here** via Harvard Dataverse. This paper was originally published in the Review of Economics and Statistics in 2019.