November 29, 2016

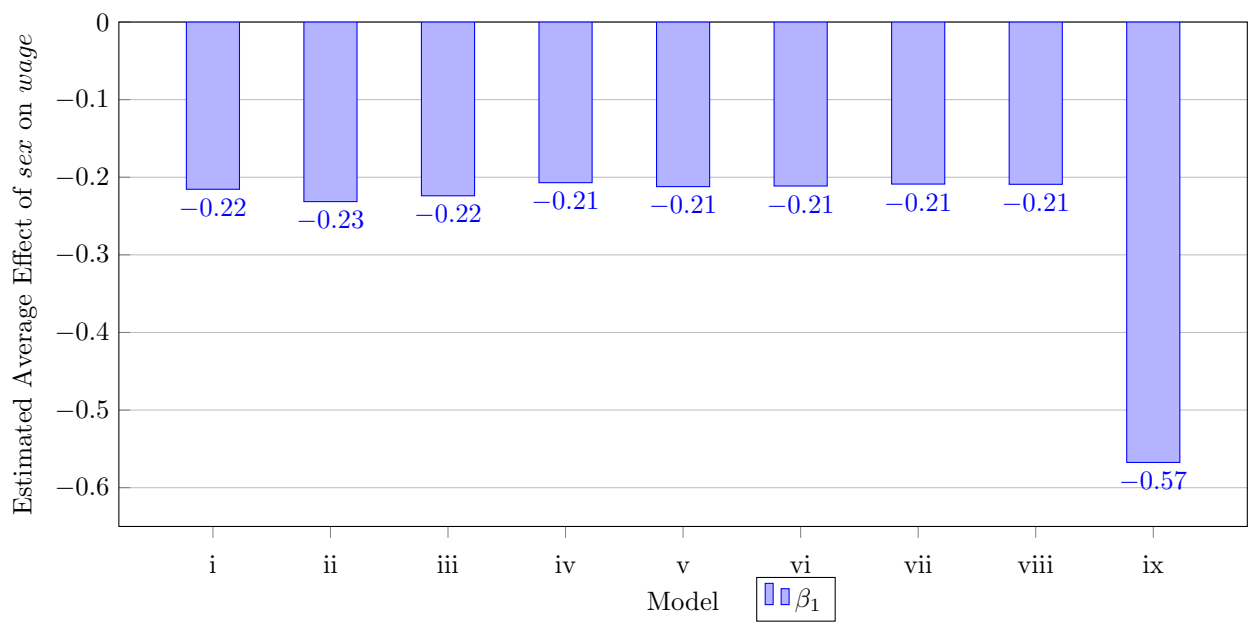
Female-male pay gap results table

Software Developers, Applications and Systems Software (occupation code 1020)

2,096 women, 8,428 men

	i.	ii.	iii.	iv.	v.	vi.	vii.	viii.	ix.
sex	-.2155***	-.2314***	-.2239***	-.2070***	-.2121***	-.2113***	-.2088***	-.2091***	-.5675***
	(.0166)	(.0161)	(.0155)	(.0154)	(.0153)	(.0153)	(.0153)	(.0153)	(.1770)
age		x	x	x	x	x	x	x	x
education ind.			x	x	x	x	x	x	x
usual work hrs.				x	x	x	x	x	x
worker class ind.					x	x	x	x	x
metro ind.						x	x	x	x
industry ind.							x	x	x
state ind.								x	x
$sex \cdot age$									x
$sex \cdot educ$									x
Adjusted R^2	0.0164	0.0930	0.1345	0.1498	0.1552	0.1556	0.1571	0.1573	0.1580
Observation count	10,524	10,524	10,524	10,524	10,524	10,524	10,524	10,524	10,524

**Note: I've chosen to round to four decimal places to further highlight changes throughout the regressions.*



MODEL i. $wage = \beta_1 sex + \varepsilon$

In this initial regression, we note there is clearly a downward trend as *sex* increases. Observe that *sex* maintains only two options, male (1) and female (2), thus it is clear that the average wage is expected to decrease from male to female.

MODEL ii. $wage = \beta_1 sex + \beta_2 age + \varepsilon$

We immediately note an impact on β_1 in the form of a negative (leftward) bias shift with the addition of the *age* variable. Without the inclusion of *age*, the average ***negative*** impact of β_1 on wage was likely underestimated in Model i when compared to this model.

MODEL iii. $wage = \beta_1 sex + \beta_2 age + \beta_3 educ + \varepsilon$

We see a notable impact on β_1 with the inclusion of *educ* in the form of positive (rightward) biase shifts, though not as significant a shift as seen in Model 2. The average ***negative*** impact of β_1 was likely overestimated Model ii when compared to this model.

MODEL iv. $wage = \beta_1 sex + \beta_2 age + \beta_3 educ + \beta_4 workhrs + \varepsilon$

With the inclusion of *workhrs* we see another shift in the estimate for β_1, a positive (rightward) bias shift. The average ***negative*** impact of β_1 on wage was likely overestimated in Model iii when compared to this model.

MODEL v. $wage = \beta_1 sex + \beta_2 age + \beta_3 educ + \beta_4 workhrs + \beta_5 class + \varepsilon$

With the inclusion of *class*, we see a negative (leftward) bias shift in β_1. The average ***negative*** impact of β_1 on wage was likely underestimated in Model iv when compared to this model.

MODEL vi. $wage = \beta_1 sex + \beta_2 age + \beta_3 educ + \beta_4 workhrs + \beta_5 class + \beta_6 metro + \varepsilon$

With the inclusion of *metro*, we see a slightly positive (rightward) bias shift in β_1. The average ***negative*** impact of β_1 on wage was likely overestimated in Model iv when compared to this model.

MODEL vii. $wage = \beta_1 sex + \beta_2 age + \beta_3 educ + \beta_4 workhrs + \beta_5 class + \beta_6 metro + \beta_7 industry + \varepsilon$

With the inclusion of *industry*, we see a slightly positive (rightward) bias shift in β_1. The average ***negative*** impact of β_1 on wage was likely overestimated in Model vi when compared to this model. This small of a shift was as expected, since we are observing a highly concentrated subset occupation of a well-defined industry.

MODEL viii. $wage = \beta_1 sex + \beta_2 age + \beta_3 educ + \beta_4 workhrs + \beta_5 class + \beta_6 metro + \beta_7 industry + \beta_8 state + \varepsilon$

With the inclusion of *state*, we see a slightly negative (leftward) bias shift in β_1. The average ***negative*** impact of β_1 on wage was likely underestimated in Model vii when compared to this model.

MODEL ix. $wage = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \underbrace{\beta_9 sex \cdot age + \beta_{10} sex \cdot educ}_{\text{my included variables}} + \varepsilon$

I chose to estimate by including two variable interactions, $sex \cdot age$ and $sex \cdot educ$, as these appeared to most significantly affect the estimate throughout the table comparison. I was interested in two factors: (1) would their inclusion further exacerbate the gender wage difference in favor of males, and (2) would their inclusion increase the R-squared value.

CONCLUSION

Indeed, we see over a 150% increase in the the average ***negative*** impact estimate of *sex* on *wage* with the inclusion of the above two interactions (β_1 changes from -0.2091 to -0.5675), resulting in a significant negative (leftward) bias shift. In other words, Models i through viii all likely underestimated the average ***negative*** impact that sex has on wages when compared to this model. While R-squared did increase, it was very little.

Appendix

```
-------------------------------------------------------------------------------
      name:  <unnamed>
       log:  D:\Users\kilo\Google Drive\Education\Classes\UCD - ECON 4811 (Fall
> 2016)\HW5.1 - Project\projlog.smcl
  log type:  smcl
 opened on:  27 Nov 2016, 19:35:10

. do "D:\Users\kilo\Google Drive\Education\Classes\UCD - ECON 4811 (Fall 2016)\m
> aster_4811.do"

. *This is the ACS 2014 .do file for the Econometrics 4811 research assignment.
.
. *Clear previous work, view and set memory.
. clear

. clear matrix

. query memory
-------------------------------------------------------------------------------
    Memory settings
      set maxvar           2048       (not settable in this version of Stata)
      set matsize          400        10-800; max. # vars in models
      set niceness         5          0-10
      set min_memory       0          0-1600g
      set max_memory       .          32m-1600g or .
      set segmentsize      32m        1m-32g

. *set virtual on
. *set mem 1000m
. set more off

.
. use "D:\Users\kilo\Google Drive\Education\Classes\UCD - ECON 4811 (Fall 2016)\
> master_dataset_4811.dta"

.
. *3,132,610 observations in the 2014 sample.
. count
  3,132,610

.
. * The following was added by me
. drop if occ2010 != 1020
(3,120,873 observations deleted)

. drop if incwage == 0
(818 observations deleted)

. drop if incwage == 999999
(0 observations deleted)

. drop if uhrswork < 35
(395 observations deleted)
```

```
. table sex

----------------------
      Sex |      Freq.
----------+-----------
     Male |      8,428
   Female |      2,096
----------------------

.
. gen lnwage = ln(incwage)

.
. reg lnwage sex, robust

Linear regression                               Number of obs     =     10,524
                                                F(1, 10522)       =     167.93
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0164
                                                Root MSE          =     .66572

------------------------------------------------------------------------------
             |               Robust
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |  -.2154826   .0166285   -12.96   0.000    -.2480775   -.1828876
       _cons |   11.59928   .0207795   558.21   0.000     11.55854    11.64001
------------------------------------------------------------------------------

. reg lnwage sex age, robust

Linear regression                               Number of obs     =     10,524
                                                F(2, 10521)       =     342.75
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0930
                                                Root MSE          =     .63931

------------------------------------------------------------------------------
             |               Robust
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |  -.2313513   .0160507   -14.41   0.000    -.2628137    -.199889
         age |   .0164707   .0006903    23.86   0.000     .0151175    .0178238
       _cons |   10.94373    .033362   328.03   0.000     10.87834    11.00913
------------------------------------------------------------------------------
```

```
. reg lnwage sex age educ, robust

Linear regression                               Number of obs     =     10,524
                                                F(3, 10520)       =     299.94
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1345
                                                Root MSE          =     .62454

------------------------------------------------------------------------------
             |               Robust
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |  -.2238923   .0155405   -14.41   0.000    -.2543547     -.19343
         age |   .0171236    .000677    25.30   0.000     .0157967    .0184506
        educ |   .0955656   .0049976    19.12   0.000     .0857694    .1053618
       _cons |   9.970441   .0670088   148.79   0.000     9.839091    10.10179
------------------------------------------------------------------------------

. reg lnwage sex age educ uhrswork, robust

Linear regression                               Number of obs     =     10,524
                                                F(4, 10519)       =     247.44
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1498
                                                Root MSE          =     .61903

------------------------------------------------------------------------------
             |               Robust
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |  -.2069715   .0154362   -13.41   0.000    -.2372295   -.1767136
         age |   .0168168   .0006678    25.18   0.000     .0155078    .0181257
        educ |   .0972345   .0049801    19.52   0.000     .0874725    .1069965
    uhrswork |   .0133306   .0010775    12.37   0.000     .0112185    .0154427
       _cons |   9.369818   .0861614   108.75   0.000     9.200925     9.53871
------------------------------------------------------------------------------
```

```
. reg lnwage sex age educ uhrswork classwkr, robust

Linear regression                               Number of obs     =     10,524
                                                F(5, 10518)       =     207.59
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1552
                                                Root MSE          =     .61709

------------------------------------------------------------------------------
             |               Robust
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |  -.2121072   .0153434   -13.82   0.000    -.2421832   -.1820313
         age |   .0170303   .0006646    25.62   0.000     .0157275    .0183331
        educ |   .0970554   .0049859    19.47   0.000      .087282    .1068288
    uhrswork |   .0139295   .0010804    12.89   0.000     .0118117    .0160473
    classwkr |   .3437876   .0801903     4.29   0.000     .1865993    .5009758
       _cons |   8.662828    .187534    46.19   0.000     8.295225     9.03043
------------------------------------------------------------------------------

. reg lnwage sex age educ uhrswork classwkr metro, robust

Linear regression                               Number of obs     =     10,524
                                                F(6, 10517)       =     176.34
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1566
                                                Root MSE          =     .61659

------------------------------------------------------------------------------
             |               Robust
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |  -.2113088   .0153321   -13.78   0.000    -.2413626   -.1812549
         age |   .0170417   .0006644    25.65   0.000     .0157393     .018344
        educ |   .0951204   .0050142    18.97   0.000     .0852915    .1049492
    uhrswork |    .013984   .0010808    12.94   0.000     .0118654    .0161026
    classwkr |    .341824   .0802233     4.26   0.000     .1845712    .4990768
       metro |   .0236822   .0057797     4.10   0.000     .0123529    .0350114
       _cons |   8.606849   .1879604    45.79   0.000     8.238411    8.975287
------------------------------------------------------------------------------
```

```
. reg lnwage sex age educ uhrswork classwkr metro ind, robust

Linear regression                               Number of obs     =     10,524
                                                F(7, 10516)       =     155.35
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1571
                                                Root MSE          =     .61644

------------------------------------------------------------------------------
             |               Robust
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |  -.2088045    .015341   -13.61   0.000    -.2388758   -.1787332
         age |   .0169733   .0006661    25.48   0.000     .0156676     .018279
        educ |   .0948653   .0050242    18.88   0.000     .0850169    .1047138
    uhrswork |   .0139252   .0010814    12.88   0.000     .0118054    .0160451
    classwkr |   .3363091   .0801722     4.19   0.000     .1791565    .4934618
       metro |   .0229403   .0058018     3.95   0.000     .0115676    .0343129
         ind |  -8.97e-06   3.35e-06    -2.68   0.007    -.0000155   -2.41e-06
       _cons |   8.684281   .1910706    45.45   0.000     8.309746    9.058815
------------------------------------------------------------------------------

. reg lnwage sex age educ uhrswork classwkr metro ind statefip, robust

Linear regression                               Number of obs     =     10,524
                                                F(8, 10515)       =     136.42
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1573
                                                Root MSE          =     .61641

------------------------------------------------------------------------------
             |               Robust
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |  -.2090987   .0153349   -13.64   0.000    -.2391579   -.1790395
         age |   .0169873   .0006663    25.49   0.000     .0156812    .0182934
        educ |   .0945984    .005017    18.86   0.000     .0847642    .1044326
    uhrswork |   .0139072   .0010827    12.84   0.000     .0117849    .0160296
    classwkr |   .3364717   .0802304     4.19   0.000     .1792049    .4937384
       metro |   .0223315   .0057996     3.85   0.000     .0109632    .0336999
         ind |  -8.79e-06   3.35e-06    -2.62   0.009    -.0000154   -2.23e-06
    statefip |  -.0004801   .0003463    -1.39   0.166    -.0011589    .0001987
       _cons |   8.700583   .1919055    45.34   0.000     8.324412    9.076755
------------------------------------------------------------------------------

.
.
. gen sa = sex*age

. gen se = sex*educ
```

```
. reg lnwage sex age educ uhrswork classwkr metro ind statefip sa se, robust

Linear regression                                      Number of obs     =     10,524
                                                       F(10, 10513)      =     113.26
                                                       Prob > F          =     0.0000
                                                       R-squared         =     0.1580
                                                       Root MSE          =      .6162

------------------------------------------------------------------------------
             |               Robust
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |   -.567496   .1770306    -3.21   0.001    -.9145096   -.2204825
         age |   .0133504   .0021905     6.09   0.000     .0090566    .0176442
        educ |   .0665543   .0159521     4.17   0.000      .035285    .0978235
    uhrswork |   .0138297   .0010823    12.78   0.000     .0117082    .0159512
    classwkr |    .336693   .0801282     4.20   0.000     .1796266    .4937594
       metro |    .022242   .0057949     3.84   0.000     .0108829    .0336011
         ind |  -8.69e-06   3.35e-06    -2.59   0.010    -.0000153   -2.11e-06
    statefip |   -.000469   .0003462    -1.35   0.176    -.0011475    .0002096
          sa |   .0030952   .0017865     1.73   0.083    -.0004067    .0065971
          se |   .0235292   .0127265     1.85   0.065    -.0014171    .0484754
       _cons |   9.127643   .2802518    32.57   0.000     8.578296     9.67699
------------------------------------------------------------------------------

.
.
end of do-file

. log close
      name:  <unnamed>
       log:  D:\Users\kilo\Google Drive\Education\Classes\UCD - ECON 4811 (Fall
> 2016)\HW5.1 - Project\projlog.smcl
  log type:  smcl
 closed on:  27 Nov 2016, 19:35:28
--------------------------------------------------------------------------------
```