Review Questions for Final - Answers

1a. Within state-year groups, calculate the age at which labor force participation is maximized or minimized. Is this a maximum or minimum? How do we know? Be careful about the definition of 'nilf" (not in labor force) when answering this question.

Differentiating the regression equation with respect to age, and setting equal to zero, yields:

```
-0.0334134 + 2 x 0.0004482 x age = 0
```

The second derivative of the regression equation with respect to age is:

```
2 x 0.0004482>0
```

Hence, the value of age that solves the first order condition will be a minimum of *nilf*, or a *maximum* of labor force participation.

Solving for **age**, we get:

age = (0.0334134)/(2 x 0.0004482) = **37.2751**

Hence, labor force participation is maximized at 37.2 years old.

1b. Within state-year groups, calculate the age at which unemployment is maximized or minimized. Is this a maximum or minimum? How do we know?

This one is slightly easier, since the question is about the dependent variable in regression 1C. The first derivative is:

```
-0.0071768 + 2 x 0.0000665 age = 0
```

And the second derivative is:

```
2 x 0.0000665 > 0
```

So, unemployment will be minimized at the point where the first order condition equals zero. This occurs when:

age = (0.0071768)/(2 x 0.0000665) = **53.9**

So, unemployment is minimized when the respondent is 53.9 years old.

1c. Going from Regression 1A to Regression 1B, some coefficients change a bit, while others do not (educ, female). What do you think the state-year fixed effects are controlling for in this case? Think omitted variables here.

The distributions of education and gender are probably reasonably equal across states and years, and therefore there is no omitted variable bias that state-years groups can correct. However, demographic groups other than gender do show reasonably different migration patterns from state to state, and this may correlate with wages. Hence, the coefficients for the demographic factors may natural changes when controlling for state-years.

1d. In Regression 1C, please interpret the coefficients on 'educ', 'female' and 'black'.

Holding age constant, within state-year groups, a one-year increase in education decreases the probability of being unemployed by 0.0187.

Holding education and age constant, within state-year groups, being a female respondent decreases the probability of being unemployed by 0.0056 relative to a male.

Holding education and age constant, within state-year groups, being a black respondent increases the probability of being unemployed by 0.064 relative to a white respondent.

---

*2a. Please interpret **precisely** the coefficient on 'female" for both regressions 2A and 2B.*

Exponentiate the coefficients to get the precise percentage change in decimal point terms:

exp(-.2545932)-1 = -0.2247682

Holding education and age constant, within years, being a female decreases wages by precisely 22.47% relative to a male.

exp(-.2543403)-1 = -0.2245442

Holding education and age constant, within state-year groups, being a female decreases wages by precisely 22.45% relative to a male.

---

*2b. Using Regression 2B, please calculate and interpret **precisely** the difference in wage for a black female compared to a white male.*

As white males are the excluded group, the difference in log real wage for black female to white male is:

lwage(bf) – lwage(wm)= -.2543-.1257 = -0.38

Expontentiating to get the precise change, we get:

exp(-0.38)-1= -0.316

So, holding education and age constant, within state-year groups, a black female earns a wage that is 31.6% less than a white male.

---

*2c. Please comment on the direction of the wage gap over time. **Precisely**, please interpret the change in the wage gap from 1983 to 2013, as evidenced in Regression 2C.*

The estimated log wage gap in 1983 is -0.334. The estimated log wage gap in 2013 is -0.203. Hence, the log wage gap has fallen by 0.131. Exponentiating, we get exp(0.131)-1 = 0.14. Hence, within state-year groups, the wage gap fell by 14% between 1983 and 2013.

2d. Suppose, that I want to test precisely the difference between the coefficient on **`female83`** *and* **`female13`**. *Please derive a regression that allows me to do this. Show your work!*

Writing female83 as f83 (and likewise for the other years), the original equation is written as:

$$\ln(rw) = \beta_{educ}educ + \beta_{age}age + \beta_{age^2}age^2 + \beta_{f83}f83 + \beta_{f88}f88 + \beta_{f93}f93 + \beta_{f98}f98 + \beta_{f03}f03 + \beta_{f08}f08 + \beta_{f13}f13$$
$$+\beta_{black}black + \beta_{hispanic}hispanic + \beta_{other}other + u$$

Defining

$$\theta = \beta_{f13} - \beta_{f83}$$
$$\Rightarrow \beta_{f83} + \theta = \beta_{f13}$$

Substituting for β_{f13}, we get

$$\ln(rw) = \beta_{educ}educ + \beta_{age}age + \beta_{age^2}age^2 + \beta_{f83}f83 + \beta_{f88}f88 + \beta_{f93}f93 + \beta_{f98}f98 + \beta_{f03}f03 + \beta_{f08}f08 + (\beta_{f83} + \theta)f13$$
$$+\beta_{black}black + \beta_{hispanic}hispanic + \beta_{other}other + u$$

$$\ln(rw) = \beta_{educ}educ + \beta_{age}age + \beta_{age^2}age^2 + \beta_{f83}(f83 + f13) + \beta_{f88}f88 + \beta_{f93}f93 + \beta_{f98}f98 + \beta_{f03}f03 + \beta_{f08}f08 + \theta f13$$
$$+\beta_{black}black + \beta_{hispanic}hispanic + \beta_{other}other + u$$

Note that you could also define $\theta_{ft} = \beta_{ft} - \beta_{f83}$ derive an equation that looks like this:

$$\ln(rw) = \beta_{educ}educ + \beta_{age}age + \beta_{age^2}age^2 + \beta_{f83}female + \theta_{f88}f88 + \theta_{f93}f93 + \theta_{f98}f98 + \theta_{f03}f03 + \theta_{f08}f08 + \theta_{f13}f13$$
$$+\beta_{black}black + \beta_{hispanic}hispanic + \beta_{other}other + u$$

where θ_{f13} is the estimated difference between 1983 and 2013, and

$$female = f83 + f88 + f93 + f98 + f03 + f08 + f13$$

Either would be acceptable answers on an exam, though the second is the easier to implement in practice.

2e. Do industries and occupations contribute to the wage gap (ie. different genders and races selecting into different industries and occupations), or is the wage gap amplified when looking within industries or occupations?

The former appears to be true. Since the wage gap falls when adding industries and occupations to our group definitions, this suggests that the industry and occupation effects that are correlated with higher wages are negatively correlated with female workers.

2f. Suppose that I claim within industry-occupation-state-year groups, the male-female wage gap is exactly twice as large as the white-black wage gap. Please write this hypothesis, and a suitable alternative. Please derive an estimating equation that allows for one to test this hypothesis.

$$H_0: \quad \frac{1}{2}\beta_{female} = \beta_{black} \quad \Rightarrow \quad \theta = \frac{1}{2}\beta_{female} - \beta_{black} = 0$$
$$H_A: \quad \theta \neq 0$$

Solving for β_{black} we get

$$\beta_{black} = \frac{1}{2}\beta_{female} - \theta$$

Substituting into the regression equation, we get:

$$\ln(rw) = \beta_{educ}educ + \beta_{age}age + \beta_{age^2}age^2 + \beta_{female}female + \beta_{black}black + \beta_{hispanic}hispanic + \beta_{other}other + u$$
$$\ln(rw) = \beta_{educ}educ + \beta_{age}age + \beta_{age^2}age^2 + \beta_{female}female + \left(\frac{1}{2}\beta_{female} - \theta\right)black + \beta_{hispanic}hispanic + \beta_{other}other + u$$
$$\ln(rw) = \beta_{educ}educ + \beta_{age}age + \beta_{age^2}age^2 + \beta_{female}\left(female + \frac{1}{2}black\right) - \theta black + \beta_{hispanic}hispanic + \beta_{other}other + u$$

2g. Write out code that does the following. Within industry-state-year groups, evaluate the differences in the male-female wage gap as a function of having a college degree. Put differently, does having a college degree affect the size/direction of the wage gap? Write out the regression specification you wish to estimate, and the code that will do it (including any variables that you need to generate).

First, we need to define a dummy variable for college-educated individuals.

```
gen college = 0
replace college =1 if educ>=16
```

Next, we need to generate an interaction between college and female

```
gen college_female = college*female
```

And finally, we estimate the following:

```
xtreg ln_rw college female college_female, fe i(state_year)
```