Problem Set 2: Due Wednesday, February 14

For this problem set, please e-mail Problem 1 and Problem 2 as separate documents.

Problem 1

Submit a 3 page outline for your Final Paper project, with 1.5-2 pages spent on each of two potential projects you will work on. The syllabus, on the bottom half of page 2, describes five questions that you should answer in your project. Do your best to address these questions for each of the two potential projects. In a couple weeks, for Problem Set 3, a classmate will look over your proposal and assess your answer to these questions, and give a short ten minute presentation based on your write-up.

Problem 2

The purpose of this exercise is to familiarize you with the Oaxaca-Blinder decomposition, a method that can be used to partition changes in the mean of a variable (e.g. workers' wages) that can be attributed to i) changes in observable differences (e.g., education, work experience) across the two points in time, versus ii) changes in the outcome conditional on these observable differences (e.g., wages conditional on education status, experience). This decomposition is a key building block of the Firpo, Fortin, and Lemieux decomposition (which tries to decompose changes over time in other distributional statistics, beside the mean) which we will discuss in Wednesday's class.

Assume, as an example, that wages at in 1980, wages are given by

$$\log w_{i,1980} = X_i \cdot \beta_{1980} + \epsilon_{i,1980}, \tag{1}$$

and that wages in 2010 are given by

$$\log w_{i,2010} = X_i \cdot \beta_{2010} + \epsilon_{i,2010} \ . \tag{2}$$

The covariates in Equation 1 and 2 include categorical variables for education—i) high school dropouts; ii) high school graduates; iii) some college; iv) college graduate; v) postgraduate education— and categorical variables of potential labor market experience—i) 0-10 years of experience; ii) 11-20 years of experience; iii) 21-30 years of experience; and iv) 31+ years of experience.

The decomposition of observed wage changes between the two periods is given begins with the following equations:

$$\begin{aligned}\Delta_O &\equiv \mathbb{E}\left[\log w_{i,2010} - \log w_{i,1980}\right] \\ &= \mathbb{E}_{2010}\left[\mathbb{E}\left[\log w_{i,2010}|X\right]\right] - \mathbb{E}_{1980}\left[\mathbb{E}\left[\log w_{i,1980}|X\right]\right] \\ &= \mathbb{E}_{2010}\left[X\right] \cdot \beta_{2010} - \mathbb{E}_{1980}\left[X\right] \cdot \beta_{1980}\end{aligned}$$

In the previous equation $\mathbb{E}_{2010}$ signifies that the expectation is over the values that the covariate could take in 2010 (and similarly for $\mathbb{E}_{1980}$ and 1980).

With additional manipulations

$$\begin{aligned}\Delta_O &= \mathbb{E}_{2010}\left[X\right] \cdot \beta_{2010} - \mathbb{E}_{1980}\left[X\right] \cdot \beta_{1980} \\ &= \mathbb{E}_{2010}\left[X\right] \cdot \beta_{2010} - \mathbb{E}_{1980}\left[X\right] \cdot \beta_{2010} + \mathbb{E}_{1980}\left[X\right] \cdot \beta_{2010} - \mathbb{E}_{1980}\left[X\right] \cdot \beta_{1980} \\ &= \underbrace{\left(\bar{X}_{2010} - \bar{X}_{1980}\right) \cdot \beta_{2010}}_{\equiv \Delta_X} + \underbrace{\bar{X}_{1980} \cdot \left(\beta_{2010} - \beta_{1980}\right)}_{\equiv \Delta_S} \end{aligned} \tag{3}$$

The "trick", here, is to add an subtract $\mathbb{E}_{1980}\left[X\right] \cdot \beta_{2010}$, the counterfactual wage that would hold if individuals with experience/education of the 1980s were paid according to wage structure in 2010. The first term in Equation 3, Δ_X, (which Fortin, Firpo, and Lemieux and others call the composition effect) reflects changes in wages due to differences in experience, education, etc... The second term, Δ_X, (the wage structure effect) reflects the part of the wage differential unexplained by compositional differences.

On the course website, I have posted a cleaned version of the MORG data set, which I produced using the replication materials from "The Contribution of the Minimum Wage to U.S. Wage Inequality over Three Decades: A Reassessment" that David Autor has posted. Make the following adjustments and variable definitions. First, construct a categorical variable that takes five values, depending on the education (use the grdatn variable in the later years of your sample and _grdhi and grdcom in the earlier years of your sample) of the worker: i) high school dropouts; ii) high school graduates; iii) some college; iv) college graduate; v) postgraduate education. Second, construct a categorical variable describing the years of potential labor market experience (equal to max[0,age-years of education - 6]): i) 0-10 years of experience; ii) 11-20 years of experience; iii) 21-30 years of experience; and iv) 31+ years of experience. In your sample, keep only full time workers (using the ft variable), who have non-missing education and experience. Finally, remove observations for workers who have wages of less than \$1 or more than \$100 (the wage variable is already in real 2012 \$).

1. Compute Δ_O, Δ_X and Δ_S. In your initial regression let the "base" education group

be workers with high school education and 11-20 years of potential experience.

2. Note that one can further break down Equation 3 as

$$
\begin{aligned}
\Delta_O &= \left(\bar{X}_{2010} - \bar{X}_{1980}\right) \cdot \beta_{2010} + \bar{X}_{1980} \cdot (\beta_{2010} - \beta_{1980}) \\
&= \left(\bar{X}_{2010,\text{education}} - \bar{X}_{1980,\text{education}}\right) \cdot \beta_{2010,\text{ education}} \\
&\quad + \left(\bar{X}_{2010,\text{experience}} - \bar{X}_{1980,\text{experience}}\right) \cdot \beta_{2010,\text{ experience}} \\
&\quad + \bar{X}_{1980,\text{education}} \cdot \left(\beta_{2010,\text{ education}} - \beta_{1980,\text{education}}\right) \\
&\quad + \bar{X}_{1980,\text{experience}} \cdot \left(\beta_{2010,\text{ experience}} - \beta_{1980,\text{experience}}\right) + \beta_{2010,0} - \beta_{1980,0}
\end{aligned}
$$

Compute the individual components of Δ_X and Δ_S.

3. Recompute your answer to part 2 with post-graduate education as the base education group. Are the magnitudes of the composition and wage structure effects robust to the omitted category?

4. Note an alternate (perhaps equally reasonable) decomposition to Equation (3) is

$$
\begin{aligned}
\Delta_O &= \mathbb{E}_{2010}[X] \cdot \beta_{2010} - \mathbb{E}_{1980}[X] \cdot \beta_{1980} \\
&= \mathbb{E}_{2010}[X] \cdot \beta_{2010} - \mathbb{E}_{2010}[X] \cdot \beta_{1980} + \mathbb{E}_{2010}[X] \cdot \beta_{1980} - \mathbb{E}_{1980}[X] \cdot \beta_{1980} \\
&= \underbrace{\left(\bar{X}_{2010} - \bar{X}_{1980}\right) \cdot \beta_{1980}}_{\equiv \Delta_X} + \underbrace{\bar{X}_{2010} \cdot (\beta_{2010} - \beta_{1980})}_{\equiv \Delta_S} \qquad (4)
\end{aligned}
$$

Recompute the Oaxaca-Blinder decomposition using Equation 4. Is the Oaxaca-Blinder decomposition robust to the base year?

As a side note: This decomposition was originally devised (in Oaxaca, 1973, "Male-Female Wage Differentials in Urban Labor Markets") to understand differences in male-female wage differences. For example using the same methodology, one may examine the extent female wages lower due to less educational attainment, or other observable differences? To address this question, we could re-label the above equations (for example with 1980 instead of male, and 2010 instead of female.)