Skills Assessment

June 11th, 2022

1 The Gender Wage Gap

What is the difference in predicted wages between men and women with the same job-relevant characteristics?

We analyze if there is a difference in the payment of men and women (*gender wage gap*). The gender wage gap may partly reflect *discrimination* against women in the labor market or may partly reflect a *selection effect*, namely that women are relatively more likely to take on occupations that pay somewhat less (for example, school teaching).

2 Data

The data set we consider is from the March Supplement of the U.S. Current Population Survey, year 2015. We select white non-hispanic individuals, aged 25 to 64 years, and working more than 35 hours per week during at least 50 weeks of the year. We exclude self-employed workers; individuals living in group quarters; individuals in the military, agricultural or private household sectors; individuals with inconsistent reports on earnings and employment status; individuals with allocated or missing information in any of the variables used in the analysis; and individuals with hourly wage below 3.

The variable of interest Y is the hourly wage rate constructed as the ratio of the annual earnings to the total number of hours worked, which is constructed in turn as the product of number of weeks worked and the usual number of hours worked per week. In our analysis, we also focus on single (never married) workers. The final sample is of size $n = 5150$.

3 Variable description

wage	Hourly wage	**ad**	Advanced Degree
lwage	Log hourly wage	**mw**	Midwest
sex	Gender (1 female) (0 male)	**so**	South
shs	Some high school	**we**	West
hsg	High school graduated	**ne**	Northeast
scl	Some college	**exp1**	Experience
clg	College graduated	**occ**	Occupational classification
		ind	Industry classification

4 Problem Set

We will construct a prediction rule for hourly wage Y, which depends linearly on job-relevant characteristics X. Consider the following log-linear regression model:

$$\log(Y) = \beta' X + \epsilon$$
$$\log(Y) = \beta_1 D + \beta_2' W + \epsilon$$

where D is the indicator of being female (1 if female and 0 otherwise) and the W's are controls explaining variation in wages. Considering transformed wages by the logarithm, we are analyzing the relative difference in the payment of men and women.

Our goals are

- Predict wages using various characteristics of workers.
- Assess the predictive performance using the (adjusted) sample MSE, the (adjusted) sample R^2, the out-of-sample MSE and R^2.

We employ two different specifications for prediction:

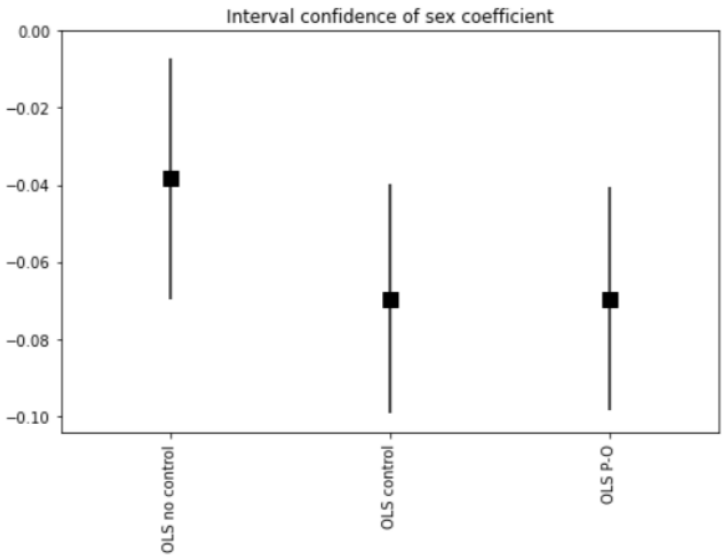
1. Basic Model: X consists of a set of raw regressors (e.g. gender, experience, education, occupation, industry and regional indicators).
2. Flexible Model: X consists of all raw regressors from the basic model plus occupation and industry indicators transformations (e.g., exp^2, exp^3, exp^4, occ^2 and ind^2) and additional two-way interactions of polynomial in experience with other regressors. An example of a regressor created through a two-way interaction is **experience** times the indicator of having a **college degree**.

   Using the *Flexible Model*, enables us to approximate the real relationship by a more complex regression model and therefore to reduce the bias. The *Flexible Model* increases the range of potential shapes of the estimated regression function. In general, flexible models often deliver good prediction accuracy but give models which are harder to interpret.

Follow the next instructions:

- Focus on the subset of college-educated workers (scl, clg variables).
- Use appropiate plots (i.e. histograms, barplots, scatter plots, pie plots, etc.) to describe main variables (wage, log-wage, sex, some high school, high school graduated, some college, college graduated, advanced degree, experience)
- Run the Basic OLS model: $lwage \sim sex + exp1 + shs + hsg + mw + so + we + occ2 + ind2$
- Run the Flexible OLS model: $lwage \sim sex + (exp1 + exp2 + exp3 + exp4 + shs + hsg + occ2 + ind2 + mw + so + we)^2$
- Compare the (adjusted) sample MSE, the (adjusted) sample R^2, the out-of-sample MSE and R^2 of both models in a table.
- Make a coefficient plot for the sex variable in both models. This is an example:

Figure 1: Coefficient Plot



You can use these references: Coefplot in R, Coefplot in Julia, Coefplot in Python

- You will also include a replication of the next figure using the data:

  Help: Use the log(*wage*) variable and the experience variable (*exp*1). You will have only two plotted lines for this College-educated workers. Do not focus on the wage gap!.

Figure 2: Wage and experience for College and High School Graduates

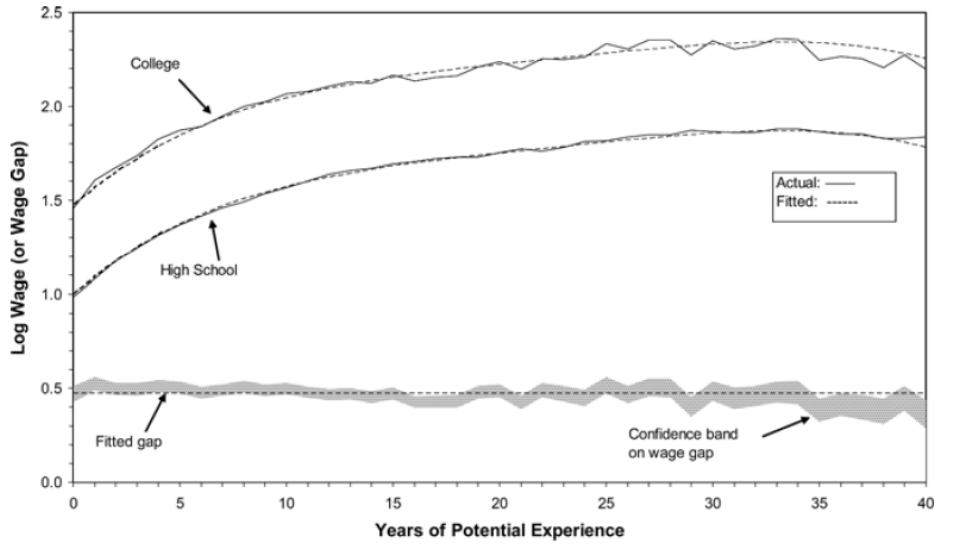


Figure 11.10. Experience Profiles and Wage Gap for College and High School Gradutes, 1989-1991 CPS.

The "Mincer equation" thirty years after schooling, experience, and earnings, Lemieux (2006: pp. 138)