The Black-white Wage Gap in the United States

Yuchen Xin, Kai Liao, Shuhan Zeng

March 23, 2021

1 Research Question

We plan to implement Heckman selection models on the National Longitudinal Surveys of Youth to solve the following problems:
1. To analyze the wage gap between the black and the white in the United States.
2. To analyze the trajectory of the wage gap between the black and the white in the United States.

2 Introduction

2.1 Motivation

Race differentials in the labor market is a persistent issue in the United States. There are substantial differences between black and white Americans in terms of unemployment rates, wage, non-wage compensation, and job characteristics. Some of these differences are not converging but rising during the past 30 years. Our research aim to identify the black-white wage gap and see how labor market discrimination evolve throughout the years.

2.2 Literature Review

Labor market discrimination is a possible reason of the black-white wage gap. However, it is difficult for economists to identify the effect of labor market discrimination and skills level separately. Theoretically, Lang and Lehmann (2011) reviews article that summarizes the economic theory behind discrimination in the labor market. Bertrand and Mullainathan (2004) use resumes with randomly assigned African-American- or White-sounding names to apply for jobs. They find that white names receive 50 percent more callbacks for interviews. Besides field experiment, economists also use observational data to identify labor market discrimination, including Neal and Johnson (1996), Fryer (2010), Altonji et al (2012), and Charles and Guryan (2008).

3 Research Design

3.1 Methodology

According to the previous literature, log hourly wages can be expressed as a function of race dummy of our interest (I_{black}) and a group of control characteristics (X) as follows:

$$logwage = X\beta + I_{black}\alpha + u \tag{1}$$

Nevertheless, without selection correction, this wage equation only refers to workers and doesn't take the people who choose not to work into account. The wage equation may introduce a selection bias when people don't make working decisions randomly. For example, due to discrimination, the black may be less likely to find a job and more likely to stay home. In this case, discrimination may influence both black people's working choice and their labor market performance, and consequently the estimated wage gap with simple OLS is biased.

In order to correct this sample selection bias, we plan to use a two-stage Heckman selection model (1979) to investigate the wage gap between the while and the black in the United States. First, we use probit models to estimate the parameters in the selection equation. With the predicted value of the selection equation, we gain a set of Inverse Mills' ratio (Z). Second, we plug the Inverse Mill's ratio obtained from the first step into the wage equation and finally get the wage gap between the black and the white. $\hat{\alpha}$ is the estimated parameter of our interest.

Selection equation:

$$workdummy = X\beta + I_{black}\alpha + v \tag{2}$$

Individual's wage is observed only if workdummy=1
Wage equation:

$$logwage = X\beta + Z\tau + I_{black}\alpha + \epsilon \tag{3}$$

After building models for the period of our interest, we will use Oaxaca Blinder decomposition (Oaxaca, 1973; Blinder, 1973) to test how much of the wage gap has been driven by the explanatory variables. Since discrimination is hard to estimate and we will not include discrimination-related variables in the model, we tend to interpret the unexplained part as the effect of discrimination.

3.2 Possible Variables

Respondent Variables:

1. Log Hourly Wage: natural log of hourly wage
2. Working dummy: =1 if the person works
3. Full-time job dummy: =1 if the person has a full-time job

Variables of our primary interests:

1. Race dummy: =1 for black or African American
2. Some interactions of race dummy and other variables like age or human capital characteristics

Control Variables:

1. Personal characteristics: marriage dummy, respondent's age, urban dummy, gender dummy, region dummy
2. Human capital characteristics: educational attainment, skill score
4. Work related characteristics:
   Cumulative number of weeks respondent has been working
   Occupation dummy
   Industry dummy
5. Household characteristics:
   Number of children under age 6 in household
   Spousal's income
   Number of persons in the family
   Family background: years of schooling of mothers and fathers

3.3 Data

National Longitudinal Surveys of Youth in 1979 (NLSY79) is a nationally representative sample of 12,686 young men and women born during the years 1957 through 1964 and living in the United States when the survey began, and therefore the survey respondents were ages 14 to 22 when first interviewed in 1979. Data are available between 1979 to 2016. It covers a bunch of variables including the age, gender, schooling, marriage status, work type, salary and etc. The reason we use this data set for our research is because it almost covers all of the personal information of the respondents, thus providing many possible factors that may influence the wage gap. For example, the wage in this survey contained details of youth and middle-aged people, thus by comparing the wage gap in ten years' time we would be able to make sure whether the wage gap is due to hourly salary or the difference in job levels among different races.

4 Outline

We plan to conduct the data work by early April, setting up the model and do the estimation work in the mid April, and finish the write-up by May 1st (we would write some of the parts as we work on the data and the model).