

Explaining the Black-White Wage Gap: Evidence from the 2015 ACS

David Chew

January 2026

Abstract

This report analyzes the Black-White wage gap in 2015 using a 5% sample of 12210 respondents from the American Community Survey (ACS). We begin by estimating baseline wage regressions to examine the relationship between education and log hourly wages, including interactions between race and education to assess differences in returns. We then apply a Oaxaca-Blinder decomposition, which incorporates education in years, age, a quadratic age term, occupation, and region to separate the raw wage gap into explained and unexplained portions.

The results show that more than half of the Black-White wage gap is explained by the aforementioned characteristics, with occupation taking on the largest chunk. While the unexplained portion is significant, it may reflect other unobserved characteristics and should be interpreted with care.

1 Data Preparation

We use the 2015 American Community Survey (ACS) to study the differences in hourly wages between black and white workers. This analysis is conducted on a 5% random sample of the overall ACS dataset.

The sample is filtered to prime-age adults, which is defined as individuals 30 to 55 inclusive, in order to focus on workers who are most connected with the labor market, minimizing those still going through school or retiring early. Furthermore, we restricted the sample to employed individuals working full time full-year, which will be defined as 30+ hours a week, and excluded the self-employed, as those earnings tend to be noisier and prone to under-reporting.

To maintain comparability with standard economic analyses, analysis is restricted to men to maintain comparability to standard labor market studies.

Hourly wages are constructed by dividing annual wage and salary income by

52 times the "usual weekly hours worked". Since these are all full year workers (50-52 weeks), this calculation works for all entries remaining in the dataset. Observations with non-positive earnings of invalid wage values were excluded. To eliminate extreme outliers and false data reporting, observations were trimmed at the 1st and 99th percentiles of the wage distribution. The primary variable of interest becomes the log of hourly wages.

Educational attainment in the ACS is reported by categories. Using information from the IPUMS codebook online, a corresponding table between the code and the number of years of education was made and applied to the column. Midpoints were used for ambiguous categorizes like "Grade 1,2,3,4" or "Grade 7 or 8".

2 Descriptive Statistics

Table 1: Descriptive Statistics by Race

race_bw	Mean log hourly wage	Mean education (years)	Mean age	N
Black	2.958	13.65	43.26	1193
White	3.254	14.14	43.19	12580

Note:

Sample restricted to employed, full-time (40+ hours) men ages 30–55 using the 2015 ACS.

Table 1 shows the mean log hourly wages, years of education, and age separately for Black and White workers. White workers earn a higher mean log hourly wage (3.254 vs. 2.958), indicating a sizable raw wage gap. There is only a small difference in the mean education attainment, with White workers having about half a year on Black workers. Average age is almost the same, suggesting that age differences are unlikely to be a major driver of the wage gap.

3 Oaxaca-Blinder Decomposition

3.1 Log Wages on race & Education

We begin the analysis by running a simple ordinary least squares (OLS) to establish baseline conditional relationships between wages, race, and education. These regressions provide a starting point for understanding how wages differ between Black and White people after controlling for observable characteristics, and show how education plays a role in explaining the wage difference in a standard regression model.

Unlike the the Oaxaca-Blinder decomposition, which focusing on decomposing the overall gap into explained and unexplained components, OLS estimates conditional wage differences under specific assumptions about how characteristics are rewarded in the labor market.

Dependent Variable:	log_hourly_wage
Model:	(1)
<i>Variables</i>	
Constant	1.892*** (0.0239)
Black	-0.2299*** (0.0162)
Education (years)	0.0959*** (0.0017)
<i>Fit statistics</i>	
R ²	0.21212
Observations	13,523

IID standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

The regression reports results from an OLS regression of log hourly wages on a black dummy indicator and years of education. The coefficient on the black indicator is negative and statistically significant, indicating that Black workers earn lower wages with the same level of education. Because the dependent variable is log hourly wages, the coefficient implies that Black workers earn approximately 23% lower wages than White workers with the same education.

The coefficient on education is positive and significant. It says that each additional year of schooling is associated with a 9.6% increase in hourly wages, on average.

Overall, the regression explains about 21 percent of the variation in log hourly wages, suggesting that while education is certainly an important determinant in wages, a considerable portion of the wage gap remains unexplained by differences in educational attainment alone.

3.2 Log Wages on race, education, and race x education

In the previous section, the black coefficient captured the average conditional wage gap assuming equal returns to education. For this next regression, we choose to include a variable of race x education allows returns to education to differ by race.

Dependent Variable:	log_hourly_wage
Model:	(1)
<i>Variables</i>	
Constant	3.243*** (0.0048)
Black	-0.2327*** (0.0164)
Education (centered)	0.0964*** (0.0017)
Black \times Education	-0.0069 (0.0067)
<i>Fit statistics</i>	
R ²	0.21218
Observations	13,523
<i>IID standard-errors in parentheses</i>	
<i>Signif. Codes: ***: 0.01, **: 0.05, *: 0.1</i>	

The OLS chart reports results from a regression of log hourly wages on race, education, and race-education interaction. Education is mean-centered so that the coefficient of the black indicator represents the black white wage gap evaluated at the average level of education in the sample.

The coefficient on education remains positive and statistically significant, indicating that more education is still associated with higher wages. The interaction term between black and education is negative, but statistically insignificant, suggesting that returns to education do not differ meaningfully between Black and White workers in the sample. As a result, the Black-White wage gap remains similar to magnitude to that in the baseline regression.

These results conclude that differences in educational attainment, rather than differences in returns to education are more likely to contribute to the wage gap. Allowing for different returns to education based on black/white does not seem to significantly alter the racial wage gap.

3.3 Twofold Oaxaca-Blinder

The Oaxaca-Blinder will decompose the Black-White wage gap into explained and unexplained components. The decomposition is estimated using years of education from the previous section, age, a quadratic age term, occupation, and geographic region. Including both age and age squared allows for a nonlinear relationship between experience and wages.

To maintain comparability across racial groups, we restrict the sample to oc-

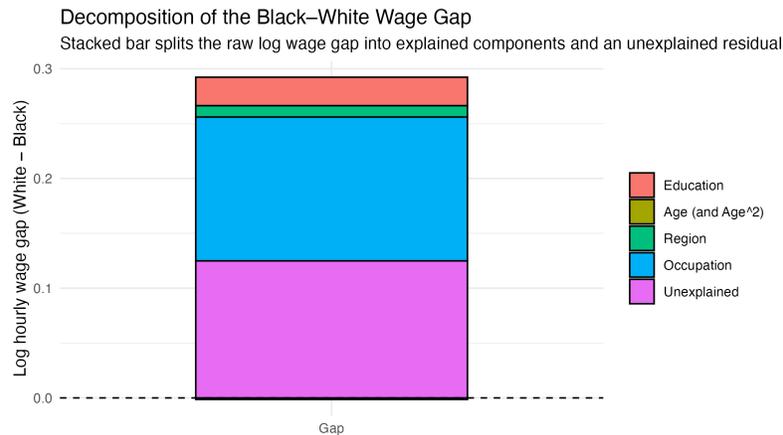
cupation categories that contain both Black and White workers. This ensures that differences in the occupational composition can be measured over different races, and won't be part of the unexplained portion.

The decomposition uses White workers as the reference group, so the explained portion of the wage gap reflects the portion of the gap that would remain if Black workers were compensated like white workers were. The unexplained component captures differences, like differential returns to education, age, occupation and location, as well as other unobserved factors.

Oaxaca–Blinder Decomposition (Twofold)

Component	Estimate
Raw gap (White – Black)	0.291
Explained	0.158
Unexplained	0.133

3.4 Daly et al. (2017) Style Graph



The figure illustrates the breakdown of the twofold Oaxaca-Blinder decomposition, showing how the raw wage gap is partitioned into explained components and the unexplained residual.

4 Discussion

A large portion of the Black-White wage gap is explained by observable characteristics, but a substantial residual remains. Of the total raw log wage gap of 0.291, approximately 54% (0.158) is explained by differences in education, age, region, and occupation, while the remaining 46% (0.133) is unexplained.

Among the explained components, occupation accounts for the largest share, meaning that racial differences in occupation have the biggest role in explaining wage disparities. This is consistent with the fact that different occupational roles pay very different. Education and region play a smaller role in the explained gap, while age is almost nonexistent.

The unexplained portion should not be interpreted as complete discrimination. There may be unobserved factors such as differences in education equality, work quality, company characteristics that are not in this dataset. Additionally, some characteristics themselves may be influenced by discrimination.