Returns to Education in the United States (2019–2024): Evidence from the CPS

Coding Sample Output (Stata)

Introduction

This report will examine the key economic outcomes in the United States for years 2019 and 2024, using data from the March Current Population Survey (CPS) downloaded via the Integrated Public Use Microdata Series (IPUMS). Specifically, the report analyzes income, earnings, educational attainment, and labor supply, as well as the conditional expectation function and the returns to education.

1. Loading data and recoding variables

I selected the CPS ASEC files for 2019 and 2024 and downloaded demographic and income variables. All analyses are person-weighted with ASECWT (appropriate for individual-level outcomes); ASECWTH is a household weight and is not used. I restricted the sample to ages 16–70 to focus on the working-age population.

Variable construction. I define the main variables as follows. Employment is a dummy equal to one if an individual reports positive annual earnings. Total annual hours are computed as the product of weeks worked last year and usual weekly hours. Real weekly and hourly earnings are constructed by dividing real total earnings by total weeks and total hours, respectively. I then take the natural logarithm of each measure to obtain log real annual, weekly, and hourly earnings.

I create a female dummy by recoding the gender variable and define four education groups: less than high school, exactly high school, some college (13–15 years), and college or more (16+ years). Finally, I recode race into four broad groups (White, Black, Asian-only, and Other) and generate a binary nonwhite indicator for simplicity.

Trimming extreme values for hours worked, weeks worked, and earnings. Table 1 below shows the 1st and 99th percentiles for `hours annual` (total annual hours worked) and `wkswork1` (weeks worked last year). Table 2 illustrates the earners falling under the 1st percentile threshold. I drop the bottom and top 1 percentiles of hours worked and the

bottom 1 percentile of low earners. I do not drop anything for weeks worked because the variable is inherently bounded.

Table 1: Trimming of Extreme Values in Hours and Weeks Worked (1%–99% Range)

Variable	Year	p1	p99	Dropped Observations
`hours_annual`	2019	184	51,948	878
`hours_annual`	2024	180	51,948	676
`wkswork1`	2019	0	52	0
`wkswork1`	2024	0	52	0

Table 2: Exclusion of Very Low Earners (Below 1st Percentile)

Year	1st Percentile Threshold ($1999)	Dropped Observations
2019	645.10	825
2024	704.54	667

Comment on quality indicators. For age, roughly 4 percent of records were allocated in both the weighted and unweighted summaries, with the remainder flagged as "no change." For sex, only about 0.15 percent was allocated, so this indicator is safe to retain. For race, approximately, over 92 percent of data remained unchanged, with the remainder being edited and replaced with a value, longitudinal carry-forward, or allocation. I retain all allocated observations. For age, race, sex, and weeks worked, over 90 percent of cases show no change, with the allocated shares for sex and weeks worked being close to zero.

Generating potential experience and restricting sample. We define potential (Mincer) experience as

$$\text{exp} \;=\; \text{age} \;-\; \text{years of education} \;-\; 6,$$

and we drop observations with experience $= 0$.

I exclude 6,833 observations in 2019 and 5,454 in 2024, where the excluded individuals are predominantly teenagers. The excluded group's median age is 18 and the 75th percentile is at 19. (see Table 3).

Table 3: Age Distribution of Observations with $exp = 0$ (Teenage Concentration)

	p1	p5	p10	p25	p50	p75	p90	p95	p99
Age (N=12,287)	16	16	16	17	18	19	20	22	22

Table 4: Summary Statistics (2019)

VARIABLES	**N**	**Mean**	**SD**	**Min–Max**
ASEC person weight	122,620	1,853.65	1,129.87	109.80 – 11,599.43
Age	122,620	42.07	15.43	16 – 70
Sex	122,620	1.52	0.50	1 – 2
Race	122,620	170.50	172.34	100 – 830
Quality flag: Age	122,620	0.14	0.63	0 – 3
Quality flag: Sex	122,620	0.00	0.12	0 – 3
Quality flag: Race	122,620	1.99	7.73	0 – 43
Usual hours/week (all jobs)	122,620	395.23	463.43	1 – 999
Educational attainment (1990 def.)	122,620	11.78	3.15	1 – 18
Quality flag: Education	122,620	0.03	0.18	0 – 1
Weeks worked last year	122,620	34.12	23.35	0 – 52
Quality flag: Weeks worked	122,620	0.01	0.11	0 – 1
Wage & salary income (prev. year)	122,620	37,356.06	64,154.59	0 – 1,599,999
Employed	122,620	0.68	0.47	0 – 1
Real total earnings (1999 dollars)	122,620	24,767.07	42,534.49	0 – 1,060,799.38
Total annual hours	87,890	7,097.64	14,129.04	4 – 51,948
Real weekly earnings (1999 dollars)	83,415	748.70	1,055.37	0.03 – 66,300
Real hourly wage (1999 dollars)	83,415	17.26	69.35	0 – 14,024.99
ln(real total earnings)	83,415	10.02	1.12	0.28 – 13.87
ln(real weekly earnings)	83,415	6.22	0.93	-3.67 – 11.10
ln(real hourly wage)	83,415	2.13	1.53	-7.36 – 9.55
Female	122,620	0.52	0.50	0 – 1
Years of education	122,620	13.57	2.74	0 – 20
Education group	122,620	2.77	1.03	1 – 4
Less than 12 years	122,620	0.13	0.34	0 – 1
12 years	122,620	0.28	0.45	0 – 1
13–15 years	122,620	0.27	0.44	0 – 1
16+ years	122,620	0.31	0.46	0 – 1
College or beyond (BA+)	122,620	0.31	0.46	0 – 1
Race (4 groups)	122,620	1.38	0.79	1 – 4
Non-white (1) vs White (0)	122,620	0.23	0.42	0 – 1

Table 5: Summary Statistics (2024)

VARIABLES	**N**	**Mean**	**SD**	**Min–Max**
ASEC person weight	97,855	2,371.76	1,446.67	124.08 – 18,536.41
Age	97,855	42.42	15.55	16 – 70
Sex	97,855	1.51	0.50	1 – 2
Race	97,855	177.30	182.08	100 – 830
Quality flag: Age	97,855	0.10	0.54	0 – 3
Quality flag: Sex	97,855	0.00	0.12	0 – 3
Quality flag: Race	97,855	2.03	7.76	0 – 43
Usual hours/week (all jobs)	97,855	383.78	460.45	1 – 999
Educational attainment (1990 def.)	97,855	11.95	3.15	1 – 18
Quality flag: Education	97,855	0.03	0.16	0 – 1
Weeks worked last year	97,855	34.53	23.16	0 – 52
Quality flag: Weeks worked	97,855	0.01	0.11	0 – 1
Wage & salary income (prev. year)	97,855	47,301.51	75,061.72	0 – 1,399,999
Employed	97,855	0.69	0.46	0 – 1
Real total earnings (1999 dollars)	97,855	25,873.93	41,058.76	0 – 765,799.44
Total annual hours	71,003	6,839.08	13,782.26	2 – 51,948
Real weekly earnings (1999 dollars)	67,413	778.07	1,055.27	0.02 – 96,272
Real hourly wage (1999 dollars)	67,413	18.08	64.40	0 – 13,675
ln(real total earnings)	67,413	10.07	1.09	0.09 – 13.55
ln(real weekly earnings)	67,413	6.27	0.92	-3.86 – 11.47
ln(real hourly wage)	67,413	2.21	1.49	-8.06 – 9.52
Female	97,855	0.51	0.50	0 – 1
Years of education	97,855	13.68	2.74	0 – 20
Education group	97,855	2.82	1.03	1 – 4
Less than 12 years	97,855	0.12	0.32	0 – 1
12 years	97,855	0.29	0.46	0 – 1
13–15 years	97,855	0.25	0.43	0 – 1
16+ years	97,855	0.34	0.47	0 – 1
College or beyond (BA+)	97,855	0.34	0.47	0 – 1
Race (4 groups)	97,855	1.41	0.82	1 – 4
Non-white (1) vs White (0)	97,855	0.24	0.43	0 – 1

Summary Statistics. The 2024 sample represents a population with more education than the 2019 sample. In 2024, on average, 34 percent of the sample has at least a bachelor's degree, compared to 31 percent in the 2019 sample. For both years, the average age is about 42 years, and slightly over half of the individuals are female, which indicates a balanced sample. Regarding race, only 23 and 24 percent of the sample represent the non-white group in 2019 and 2024, respectively. Real total earnings and real hourly wages are higher in 2024 than in 2019, revealing an increase in earnings post-pandemic.

The 2024 sample has fewer observations than the 2019 sample; however, it remains sufficiently large. The sample size is smaller for total annual hours, real weekly earnings, real hourly wages, and their logarithmic transformations in both years because they are restricted to employed individuals only.

2. The distribution of earnings, wages, and hours

We compare the distributions of annual earnings, hourly wages, and annual hours for 2019 and 2024. Tables 6 and 7 report the *mean*, *median*, *25th*, and *75th* percentiles of these variables, both unweighted and weighted. Comparing the data from 2019 to 2024, we see higher annual earnings and hourly wage at every quartile in 2024. In contrast, we see higher lower hours worked in the 75th percentile in 2024, while the median remains the same in both years. Additionally, we see similar results for both weighted and unweighted statistics, which indicates modest differences between weighted and unweighted sample composition.

Table 6: Distributional Summaries (2019): Unweighted vs. Weighted

	Unweighted				**Weighted (asecwt)**			
Variable	**Mean**	**Median**	**p25**	**p75**	**Mean**	**Median**	**p25**	**p75**
Annual Real Earnings (1999 $)	26,271.91	16,575.00	0.00	36,465.00	26,297.61	16,575.00	0.00	36,465.00
Annual Hours Worked	7,130.07	2,080.00	2,080.00	2,600.00	6,982.68	2,080.00	2,080.00	2,600.00
Hourly Wage (1999 $)	16.99	12.75	7.01	20.72	17.11	12.75	7.08	20.72
Observations				114,418				weighted by `asecwt`

Notes: Unweighted columns use equal weights. Weighted columns use `asecwt`.

Table 7: Distributional Summaries (2024): Unweighted vs. Weighted

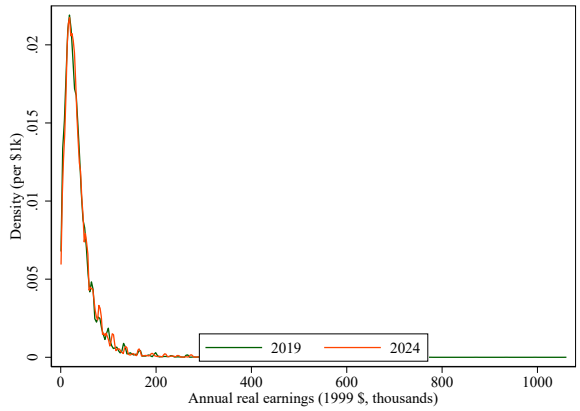
	Unweighted				**Weighted (asecwt)**			
Variable	**Mean**	**Median**	**p25**	**p75**	**Mean**	**Median**	**p25**	**p75**
Annual Real Earnings (1999 $)	27,443.21	18,598.00	0.00	38,290.00	27,303.12	18,051.00	0.00	38,290.00
Annual Hours Worked	6,848.39	2,080.00	2,080.00	2,400.00	6,738.96	2,080.00	2,080.00	2,340.00
Hourly Wage (1999 $)	17.82	13.15	7.89	21.56	17.92	13.15	7.89	21.65
Observations				91,317				same sample; weighted by `asecwt`

Notes: Unweighted columns use equal weights. Weighted columns use person weights `asecwt`.

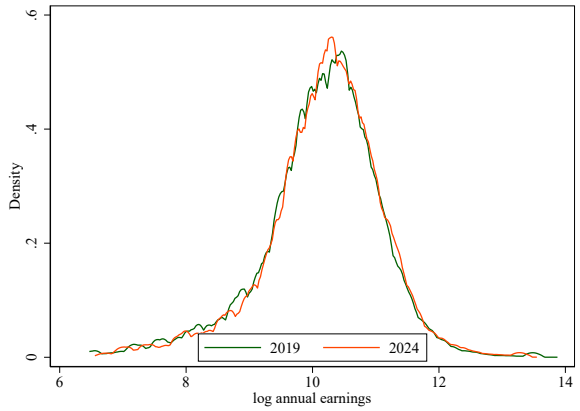
Figure 1 displays distributions of real annual earnings, hourly wages, and hours, as well as of their logarithmic transformations. Distributions in levels show significant right-tail out-liers, however, logarithmic transformations reduce skewness.

Figure 2 exhibits the conditional distribution of log hourly wages in 2024 by gender and by white vs. non-white. The gender distribution is skewed to the left and shows modest differences between male and female, with the distribution representing female log hourly wage being shifted slightly to the left, exhibiting, on average, lower wages. In the figure illustrat-ing log hourly wage conditioned by race, the distributions are almost the same. However, it is important to note that my variable white versus non-white is too broad to represent differences in economic outcomes conditioned by race.

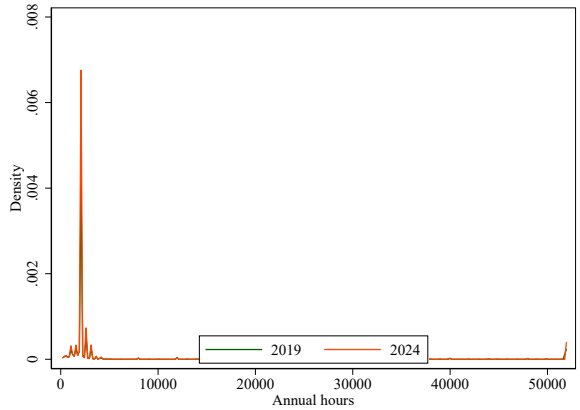
Finally, when looking at the standard deviation and different percentiles in log hourly wages, we see that pay differences shrank overall from p90-p10 equal to 3.688 in 2019 to 3.530 in 2024. This was primarily driven by the compression of the lower tail (p50–p10: 2.732 decreased to 2.566) , while the upper tail remained relatively unchanged (p90–p50: 0.956 decreased to 0.965). This shows that low-wage workers experienced wage growth from 2019 to 2024.



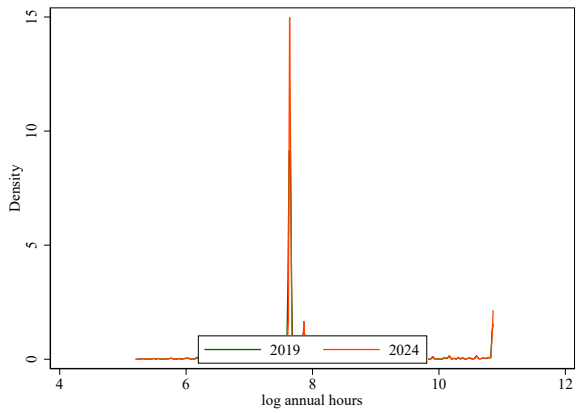
(a) Annual earnings (levels)



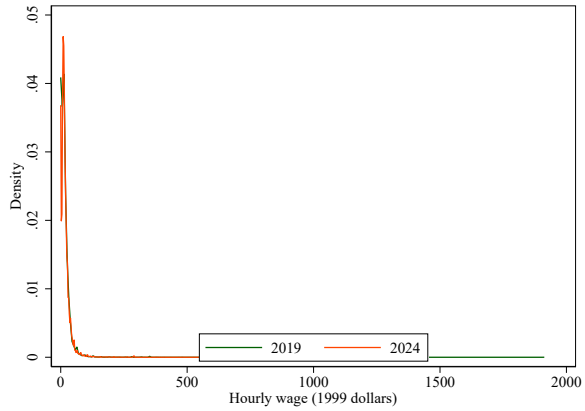
(b) Log annual earnings.



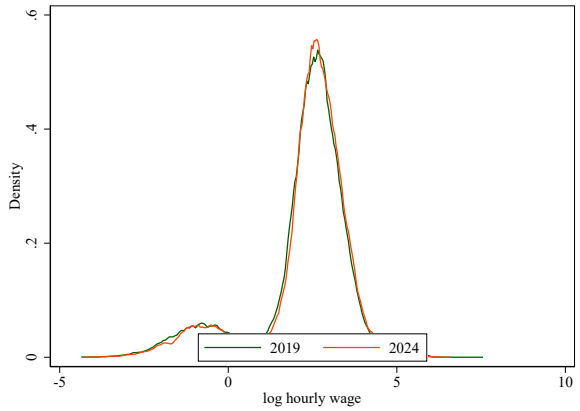
(c) Annual hours (levels)



(d) Log annual hours (2019 vs 2024).

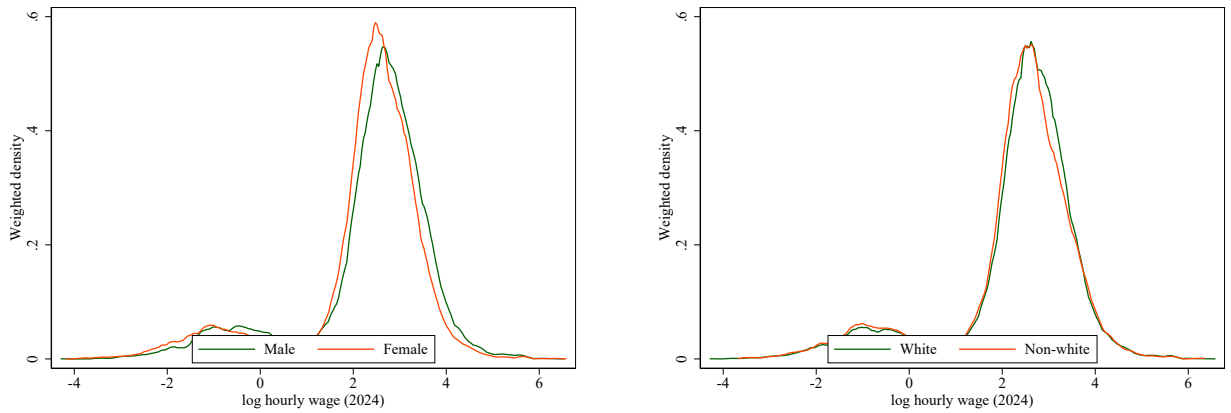


(e) Hourly wages (levels)



(f) Log hourly wages (2019 vs 2024).

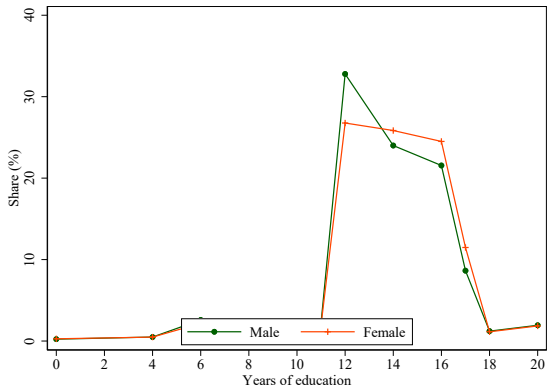
Figure 1: Kernel density plots of earnings, hours, wages, and their logarithmic transformations (2019 vs 2024). Units in thousands of 1999 dollars.



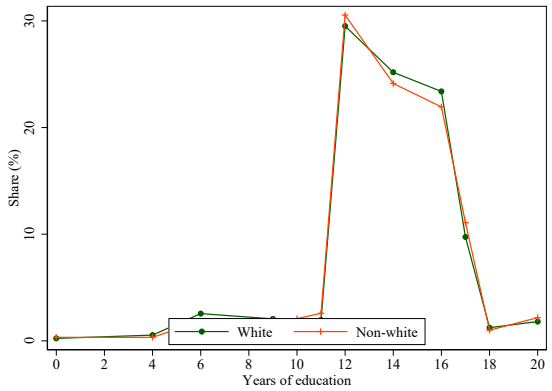
(a) By gender (b) By race

Figure 2: Weighted kernel densities of log hourly wages in 2024.

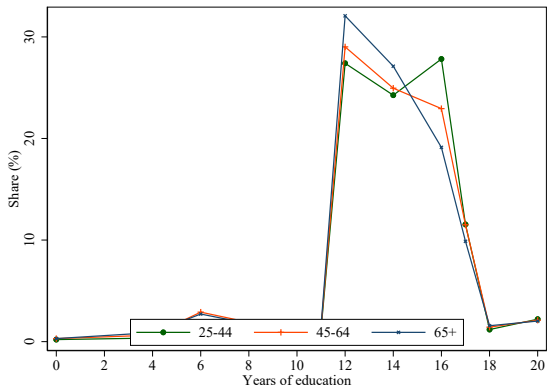
Figure 3 exhibits weighted distributions of years of education by age group, gender, and race. People aged 24 to 44 have the highest share of at least bachelor's degree group and the lowest high school only share. The age group 65+ shows the exact reverse: they display the highest shares high school only and the lowest in at least bachelor's. The 45-64 group lies in between. This shows strong trends in educational attainment: younger population chooses to stay in school longer than previous generations. By gender, we see that more men end their education after high school, and more women attain bachelor's degrees. Finally, the white versus non-white comparison shows a slightly higher share with university degrees among the white group.



(a) Weighted distribution by sex (2024).



(b) Weighted distribution by race (2024).



(c) Weighted distribution by age group (2024).

Figure 3: Weighted distributions of years of education (CPS ASEC 2024).

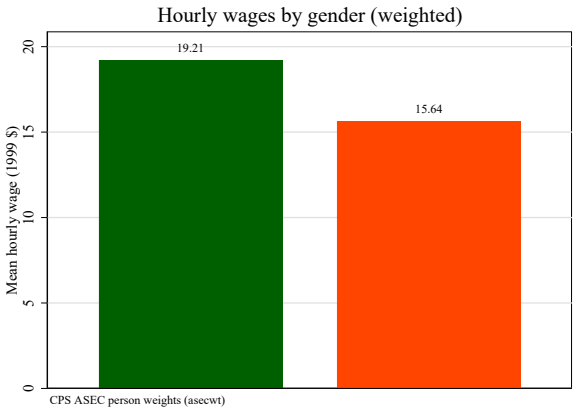
The Wage Structure and Conditional Earnings Function.

Figure 4 compares annual earnings and hourly wages by gender, race, and age group. When looking at gender, we see that men out-earn women in terms of both annual earnings and hourly wages: 32,718 vs 20,980 dollars (annual) and 19.21 vs 15.64 dollars (hourly). By race, the figures exhibit the Asian group has the highest earnings and wages, 35,071 and 23,43 dollars, respectively, followed by the White, Black, and Other groups in that order. The graphs by age group show that the highest values occur for the middle group (45–64), with annual earnings dropping significantly for the oldest age group (65+).

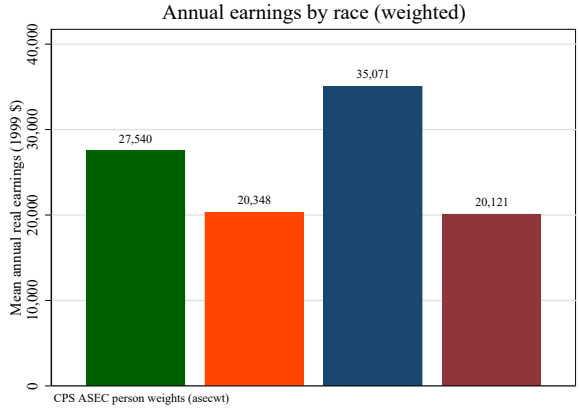
Figure 5 shows Mean log hourly wage by education group for men and women. For both genders log hourly wages rise with a further education group, however, there is a strong gender gap within each group. Men's wages exceed women's at every education level, with the largest gap being in less than 12 years of education group.

Figure 6 exhibits the conditional expectation function (CEF) of log real earnings and log hourly wage conditional by years of education in 2024. Log annual earnings and hourly wage is relatively before finishing 12 years of schooling and start to rise more notably once the individual graduates from high school, with the highest expected log annual earnings and log hourly wage at 18 years of education. Figure 7 compares CEF of log hourly earnings for 2019 and 2024. Both years show the same sharp increase starting at 12 years of education which plateaus at 18 years.

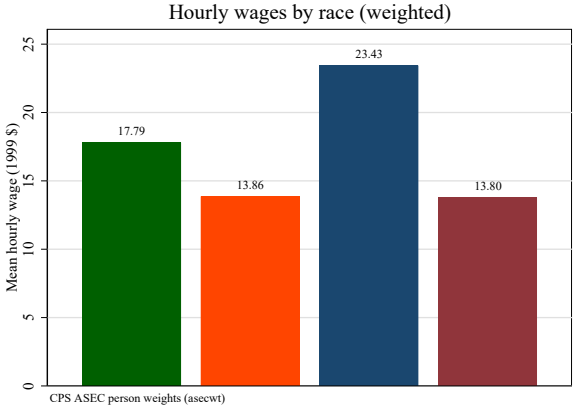
(a) Earnings by gender



(b) Wages by gender

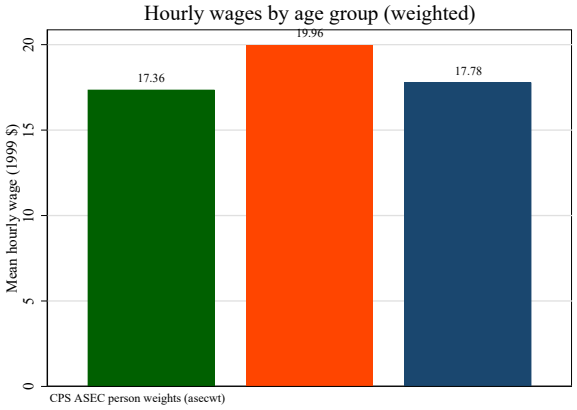


(c) Earnings by race



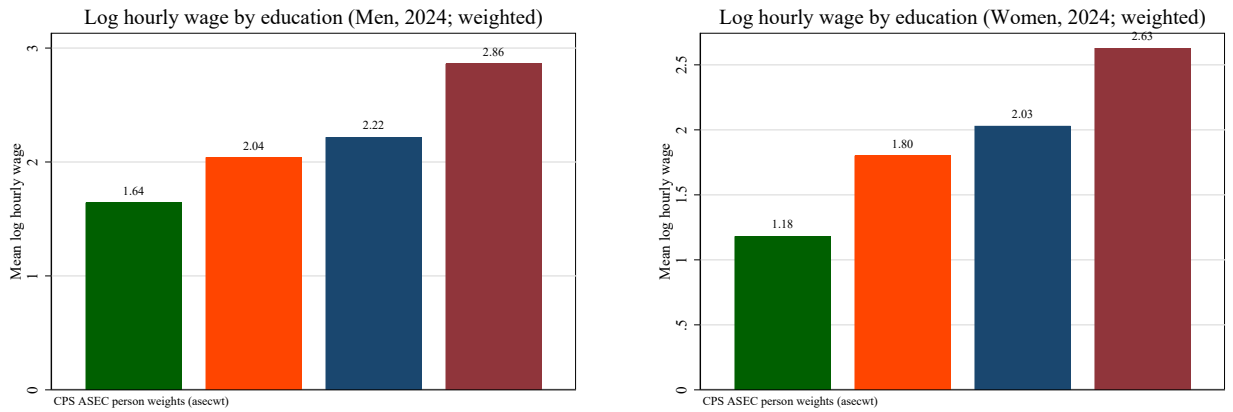
(d) Wages by race

(e) Earnings by age group



(f) Wages by age group

Figure 4: Weighted means of annual earnings and hourly wages by gender, race, and age group. Legend. *Gender:* Male (left bar, green), Female (right bar, orange). *Race:* White, Black, Asian-only, Other (bars shown left-to-right). *Age groups:* 25–44, 45–64, 65+ (bars shown left-to-right).



(a) Men (2024) (b) Women (2024)

Figure 5: Mean log hourly wage by education group (weighted).
Legend: less than 12 =12 (HS) 13–15 16+ (BA+)

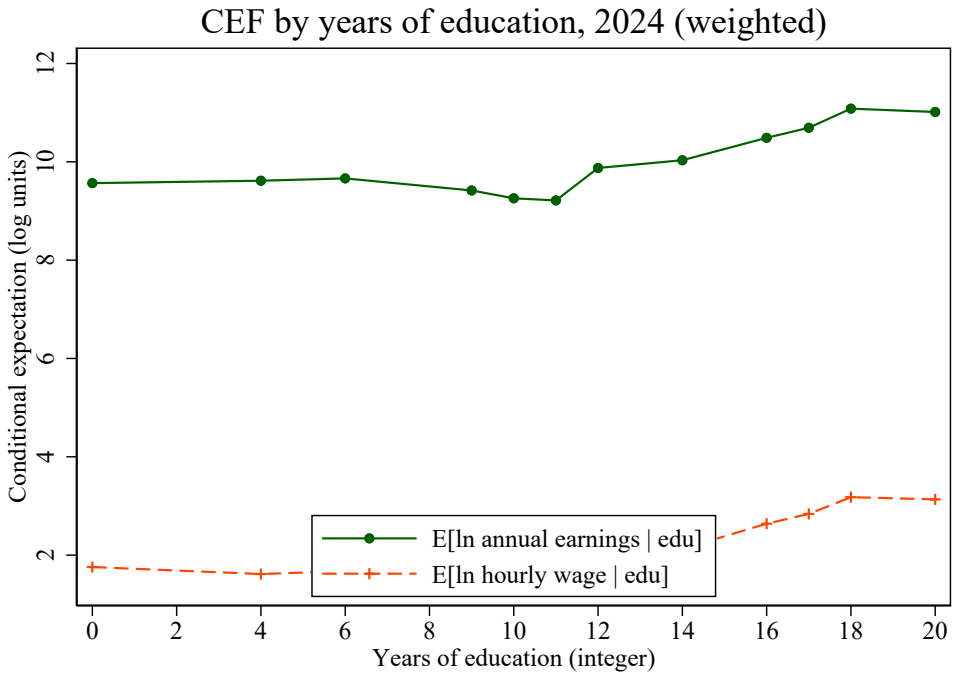


Figure 6: Conditional expectation functions in 2024 (weighted): $E[\ln \text{earnings} \mid \text{edu}]$ vs. $E[\ln \text{hourly wage} \mid \text{edu}]$.

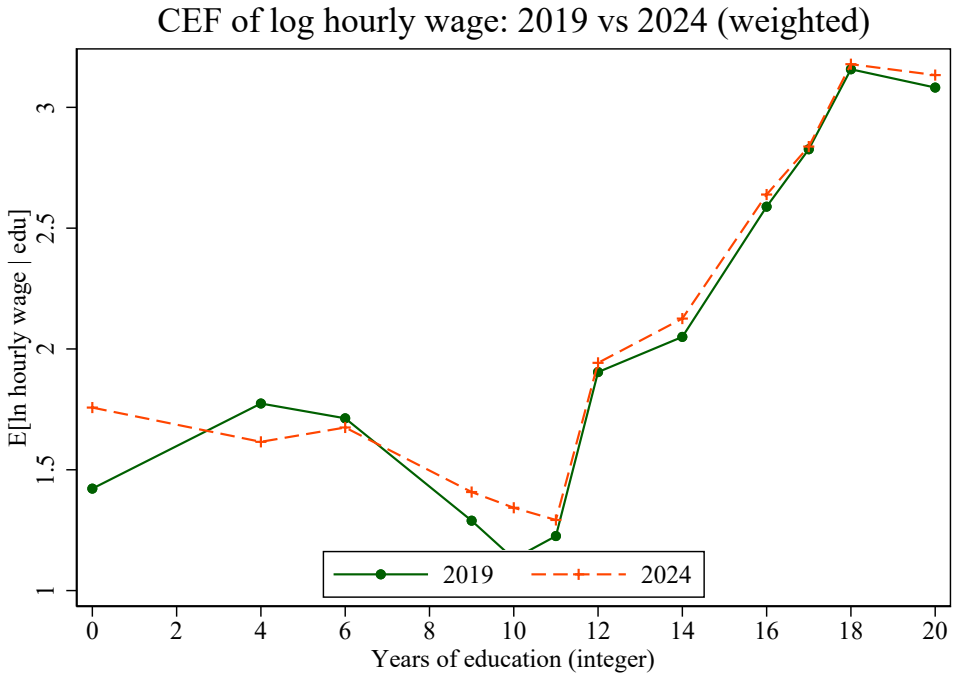


Figure 7: $E[\ln \text{hourly wage} \mid \text{edu}]$ in 2019 vs. 2024 (weighted).

The OLS returns to education.

Using the 2024 sample, I run Mincer regression model to estimate returns to education. The dependent variable is ln(hourly wage); the key regressor is years of schooling. I add sequentially controls for gender and race, as well as polynomials in potential experience.

Main specification

$$\ln w_i = \beta_0 + \beta_1 \text{educ}_i + \gamma_1 \text{exp}_i + \gamma_2 \text{exp}_i^2 + \gamma_3 \text{exp}_i^3 + \gamma_4 \text{exp}_i^4 + \delta' X_i + \varepsilon_i,$$

where X_i includes gender and race. Return to education coefficient shows similar and statistically significant effect across the three specifications in Table 8. In the main specification for 2024, the coefficient on years of schooling is 0.145, implying about 15 percent higher hourly wage per additional year of education. The model does not comprehensively explain returns to education as overall fit is modest (R squared close to 0.12). If an individual is a woman, the hourly wage is lower by almost 23 percent. When an individual is not white, the hourly wage is lower by about 6 percent. Tables 9 and 10 show return to education by race and gender. Returns are the strongest for Black workers (0.164) followed by Asian (0.148) and Other (0.146). By gender, women benefit more from an additional year in school (0.156) compared to men (0.136). Education yields large premiums across groups; however, gender and race differences in hourly wages persist.

Table 8: OLS Returns to Education (Dependent: ln Hourly Wage), 2024

	(1)	(2)	(3) Main
Years of education (from EDUC99)	0.141***	0.146***	0.145***
	(0.002)	(0.002)	(0.002)
Potential experience (age − educ_years − 6)	0.062***	0.062***	0.132***
	(0.002)	(0.002)	(0.008)
$\exp^2$	−0.001***	−0.001***	−0.006***
	(0.000)	(0.000)	(0.001)
$\exp^3$			0.000***
			(0.000)
$\exp^4$			−0.000***
			(0.000)
Female (1)		−0.230***	−0.229***
		(0.010)	(0.010)
Non-white (1)		−0.063***	−0.063***
		(0.012)	(0.012)
Observations	63,954	63,954	63,954
R-squared	0.109	0.116	0.119
RMSE	1.298	1.293	1.291

Notes: Robust standard errors in parentheses. (1) educ + exp + $\exp^2$; (2) + gender, non-white; (3) + $\exp^3$, $\exp^4$ (main spec). *** $p < 0.01$.

Table 9: OLS Return to Education by Race (2024)

	White (1)	Black (2)	Asian-only (3)	Other (4)
Years of education (from EDUC99)	0.138***	0.164***	0.148***	0.146***
	(0.002)	(0.007)	(0.008)	(0.012)
Observations	49,258	6,942	4,887	2,867
R-squared	0.117	0.106	0.119	0.105
RMSE	1.287	1.286	1.224	1.420

Notes: Robust SE in parentheses; each column is a separate regression with the same controls used in Stata. *** $p < 0.01$.

Table 10: OLS Return to Education by Gender (2024)

	Male (1)	Female (2)
Years of education (from EDUC99)	0.136***	0.156***
	(0.003)	(0.003)
Observations	33,354	30,600
R-squared	0.124	0.110
RMSE	1.284	1.297

Notes: Robust SE in parentheses; each column is a separate regression with the same controls as your gender runs. *** $p < 0.01$.

Figure 8 shows the fitted CEF for log hourly wages with 95 percent bands. The fitted CEF rises almost linearly with years of education, showing a small kink around 12 years (high school completion). In table 11, I used log annual and weekly earnings and I still reach similar conclusions as in Table 8, where I used log hourly wage: each additional year of education has a statistically significant and positive relationship with earnings, while the gender and race discrepancies persist. One additional year of education is associated with a 14 and 13 percent increase in annual and weekly earnings, respectively.

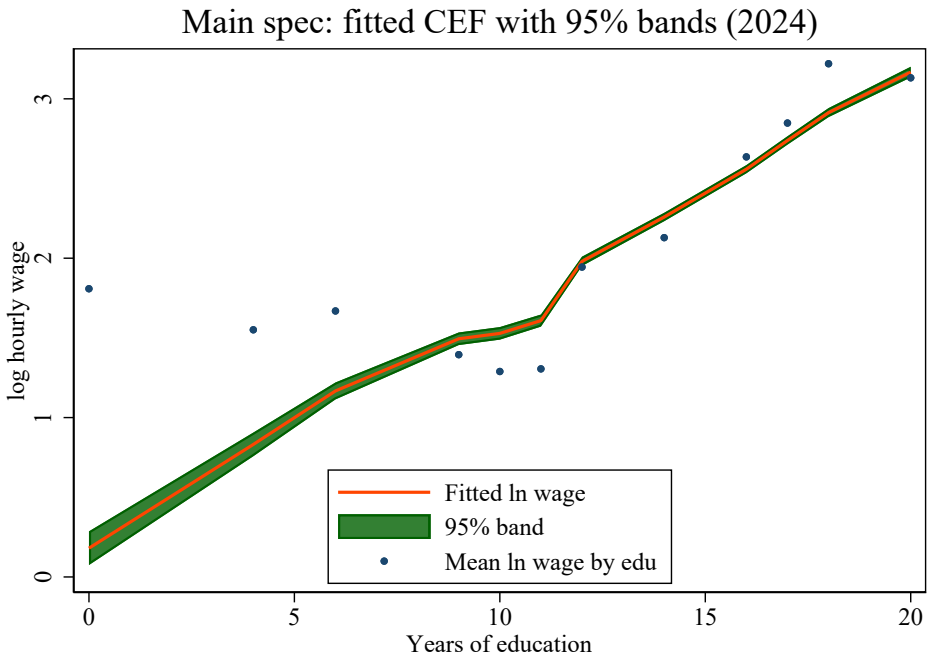


Figure 8: Main-spec CEF of log hourly wages with 95% bands (2024; weighted).

Table 11: Main spec: ln Annual vs. ln Weekly earnings (2024)

(1) ln_annual	(2) ln_week	
Years of education (from EDUC99)	0.136***	0.129***
	(0.001)	(0.001)
Potential experience (age - educ_years - 6)	0.162***	0.112***
	(0.005)	(0.004)
exp2	-0.008***	-0.005***
	(0.000)	(0.000)
exp3	0.000***	0.000***
	(0.000)	(0.000)
exp4	-0.000***	-0.000***
	(0.000)	(0.000)
Female	-0.367***	-0.344***
	(0.006)	(0.006)
Non-white	-0.051***	-0.039***
	(0.007)	(0.007)
Obs.	63954	63954
R^2	0.271	0.267
RMSE	0.784	0.701

Table 12, in contrast to the main specification which utilized years of education, uses four education groups. Because I use groups, the premiums are discrete, and therefore, larger. Relative to less than 12 years, hourly wages are about 49 percent higher for graduating high school, 81 percent higher for 13–15 years, and 228 higher for BA+.

Table 12: Returns to education groups (base: $<$ 12 years), 2024

(1) ln_wage	
=12 years (HS/GED)	0.399***
	(0.026)
13–15 years	0.595***
	(0.027)
16+ years (BA+)	1.189***
	(0.026)
Potential experience (age - educ_years - 6)	0.135***
	(0.008)
exp2	-0.006***
	(0.001)
exp3	0.000***
	(0.000)
exp4	-0.000***
	(0.000)
Female	-0.232***
	(0.010)
Non-white	-0.056***
	(0.012)
Obs.	63,954
R^2	0.126
RMSE	1.286