Global Academic Frontiers, Vol. 3, Issue 3, 2025, pp. 1-13 Print ISSN 2995-5688 Online ISSN 2995-570X
DOI: https://doi.org/10.5281/zenodo.17018216

Understanding Wage Disparities Through Educational Attainment

Jiayi Zhang [1], Xiaoran Han [1*]

[1] College of Art and Sciences, Boston University, Boston, MA, USA
*Corresponding author Email: chrishan0318@163.com
Received 20 May 2025; Accepted 11 July 2025; Published 1 September 2025

Abstract: This paper investigates the causal effect of education on wages using data from the 2018 American Community Survey (ACS) accessed through IPUMS. Our paper tests the causal effect of educational attainment on wages using individual-level data by using quarter of birth as an instrumental variable for education. We hypothesize that higher education leads to significantly higher wages. Based on our IV regression model, we find that each additional year of education is associated with approximately a 20.8% increase in log wages. These results support the importance of education in explaining wage differences, although other factors such as gender also play a notable role.

Keywords: Education, Wage Disparity, Instrumental Variables, American Community Survey, Labor Economics

I. Introduction

The relationship between educational attainment and labor market outcomes has been a central focus in labor economics. According to the foundational theory of human capital proposed by Becker (1964), individuals invest in education to increase their productivity, which in turn leads to higher earnings. Mincer (1974) further formalized this relationship through a widely adopted earnings function that regresses log wages on years of schooling and labor market experience. While the theoretical linkage is intuitive, empirical estimation of the causal effect of education on earnings is complicated by endogeneity concerns. Unobserved factors such as innate ability, family background, or motivation may influence both educational attainment and labor market success, biasing ordinary least squares (OLS) estimates.

To address this challenge, we adopt an instrumental variables (IV) approach using quarter of birth as an instrument for education, following the methodology of Angrist and Krueger (1991). The quarter-of-birth instrument exploits the exogenous variation created by compulsory schooling laws and school entry age cutoffs, which affect how long individuals stay in school without being correlated with innate ability or motivation. This approach allows us to better isolate the true causal impact of education on wages.

We use microdata from the 2018 American Community Survey (ACS) accessed via IPUMS to estimate the effect of educational attainment on log wages. To further reduce omitted variable bias, we include a range of controls such as age, gender, race, and hours worked per week. In our IV regression model, we find that each additional year of education is associated with a significant increase in log wages, suggesting a strong causal relationship. This finding contributes to the growing empirical literature on wage determination and underscores the policy relevance of education as a lever for income mobility and labor market equity. Moreover, we discuss how gender and other demographic factors intersect with education in explaining wage disparities.

II. Literature Review

A substantial body of literature has documented a positive relationship between education and earnings. Early empirical studies, including those by Mincer (1974) and Becker (1964), posited that each additional year of schooling increases an individual's productivity and thus their market wage. These models laid the foundation for estimating the returns to education using regression-based approaches. More recently, Card (1999) conducted a comprehensive review of the empirical evidence and concluded that the returns to education are both economically and statistically significant, though estimates vary depending on methods and data sources.

However, a key concern in estimating the return to education is endogeneity bias. Individuals with higher innate ability or better socioeconomic backgrounds may both attain more education and earn higher wages, leading to an upward bias in OLS estimates. To address this, several studies have turned to instrumental variables. Among the most influential is the work of Angrist and Krueger (1991), who used quarter of birth as an instrument for years of education. They argued that individuals born earlier in the year typically start school later and are therefore more likely to leave school earlier, leading to small but meaningful differences in educational attainment. Their IV estimates of the return to schooling—ranging from 8% to 13% per additional year—were higher than their OLS counterparts, emphasizing the importance of correcting for endogeneity.

Complementary research by Oreopoulos (2006) used changes in compulsory schooling laws in the United Kingdom and Canada as instruments and found large and robust returns to education. He also showed that these returns were particularly high for individuals from disadvantaged backgrounds, reinforcing the role of education in promoting economic mobility. Similarly, Sorel and Shinners (2019) analyzed data from Georgia using multiple regression models and found that each level of educational attainment was associated with a 12.6% increase in wages in a simple linear model, but the effect decreased to 5.7% after including demographic controls—highlighting the influence of confounding variables such as gender and race.

Furthermore, Heckman, Lochner, and Todd (2006) emphasized that while education is a strong predictor of earnings, non-cognitive skills, early childhood investments, and family environments also significantly contribute to labor market success. These findings suggest that policies aimed solely at increasing educational attainment may not fully address wage disparities unless they also consider broader social and economic factors.

Taken together, the literature demonstrates a consistent and substantial return to education, though the magnitude of the effect depends critically on the estimation strategy. Our study contributes to this ongoing discourse by applying a well-established IV methodology to recent ACS data, thereby providing updated evidence on the causal impact of education on earnings in the United States.

III. Data

A. Source of Data

The analysis uses data from the 2018 American Community Survey (ACS), accessed through IPUMS. The ACS is a nationally representative annual survey conducted by the U.S. Census Bureau, designed to collect detailed demographic, social, economic, and housing information. For our analysis, we use a 1% sample of the 2018 ACS and focus on variables relevant to wage determination, including age, sex, race, educational attainment (both general and detailed versions), usual hours worked per week, and wage and salary income. We also include the individual's quarter of birth as an instrumental variable to address potential endogeneity in education. This rich microdata allows for robust regression analysis of the relationship between educational attainment and earnings.

Table 1. Summary of the Variables

Variable	Obs	Mean	Std. Dev.	Min	Max

incwage	1,548,402	52,613.331	65,699.461	4	718,000
educyrs	1,548,402	14.000	2.955	0	21
age	1,548,402	43.18	15.061	18	96
female	1,548,402	0.483	0.500	0	1
uhrswork	1,548,402	38.774	12.567	1	99
NonWhite	1,548,402	0.798	0.402	0	1

To supplement the descriptive statistics table, we highlight a few key observations about the minimum and maximum values of our dependent variable, income (incwage), and key regressor, years of education (educyrs). The raw wage data ranges from $4 to $718,000, with a mean of $52,613.33 and a standard deviation of $65,699.46. This large variation and extreme upper bound suggest the presence of outliers or a skewed distribution, which is why we later transform income into a natural logarithm (lnwage)to improve interpretability and reduce the influence of extreme values. Similarly, educyrs ranges from 0 to 21, corresponding to the full educational spectrum from no schooling to doctoral degrees, as classified by the IPUMS detailed education codes. We excluded records with education codes labeled as "N/A" or "missing" and mapped each valid category into equivalent years of schooling to construct a continuous variable. The minimum of 0 reflects individuals with no formal education or only reached kindergarten, while the maximum of 21 corresponds to those holding doctoral degrees. Additionally, the age variable ranges from 18 to 96, with a mean of 43.18, ensuring our sample includes only working-age adults. These cleaned and transformed variables provide a more reliable basis for the regression analysis and ensure that extreme or non-informative values do not distort our results.

B. Models and Results

The scatter plot in Figure 1 depicts the relationship between educational attainment(educyrs) and the natural logarithm of wage (lnwage). Each dot represents an observation, and the red line represents the fitted values from a simple linear regression. From the plot, we observe a generally positive trend: as education level increases, the log of wages tends to increase as well. This supports the human capital theory that higher education leads to higher earnings. However, the relationship is not perfectly linear—there is considerable variation in wages within each education level, especially at lower education levels. Despite this dispersion, the upward slope of the fitted line indicates that on average, each additional level of education is associated with higher logged wages. The fitted line provides a reasonable linear approximation of the underlying pattern, justifying the use of a linear regression model for this analysis.

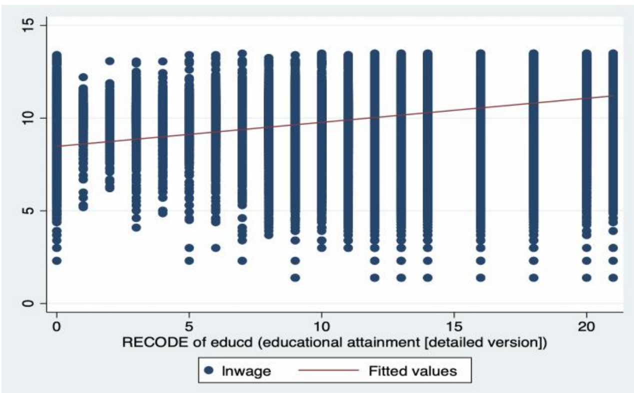

Figure 1. Scatter Plot of Logged-Wage versus Education

IV. Econometric Analysis

A. Simple Regression Model Selection

In both models, the slope coefficient on *educyrs* is highly statistically significant ($p < 0.001$), indicating a strong relationship between education and wages. The positive coefficients indicate that more years of education are associated with higher earnings, aligning with human capital theory. In the raw income model, the coefficient of 6984.07 means that each additional year of education is associated with an average increase of about $6,984 in annual income, holding other factors constant. In the log-linear model, the coefficient of 0.1252 suggests that a 1-year increase in education is associated with approximately a 12.52% increase in wages, interpreting it as a semi-elasticity since the dependent variable is in logs but the regressor is in levels. Psacharopoulos and Patrinos (2018) provided extensive cross-country evidence showing that returns to education remain consistently high, particularly in developing countries, and that each additional year of schooling significantly boosts income. Their findings reinforce the reliability of our result and its policy relevance.

Both models have relatively low R-squared values, though the log-linear model is slightly lower (0.0883 vs. 0.0996). However, after comparing the distribution of the raw wage variable (*incwage*) to its logarithmic transformation (*lnwage*), Figure 2 presents the histogram of raw wages, which is highly right skewed with a long tail extending toward higher income values.

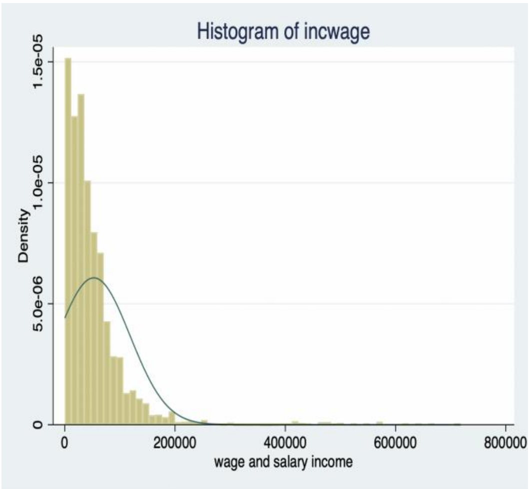

Figure 2. Histogram of Unlogged Wage

The distribution exhibits extreme values, with the maximum reaching $718,000, and a large proportion of observations concentrated at the lower end. Such skewness violates the normality assumption underpinning classical linear regression models and can lead to inefficient and biased estimates. Biewen and Fitzenberger (2005) emphasized that log-transforming skewed wage variables helps achieve a closer approximation to normality and homoscedasticity, leading to more efficient estimation in earnings regressions.

In contrast, Figure 3 displays the histogram of logged wages (lnwage).

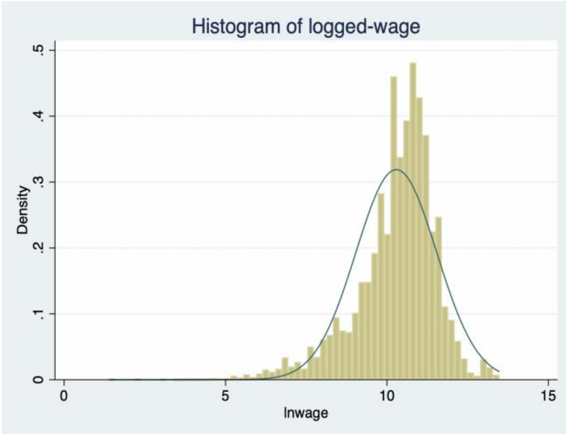

Figure 3. Histogram of Logged Wage

After log transformation, the distribution appears significantly more symmetric and bell-shaped, closely resembling the normal distribution. The transformation compresses the scale of wages and reduces the influence of outliers, resulting in a more homoscedastic variance structure. Additionally, using the log of wage facilitates elasticity-based interpretations of regression coefficients, which are common in labor economics research. Based on this comparison, we adopt the logged wage (lnwage) as our dependent variable in all subsequent regression analyses. Therefore, our equation of the simple regression model is shown as below:

$$\log(wage) = \beta 0 + \beta 1 \cdot educyrs + u, \quad (1)$$

$$\log(wage) = 8.4734 + 0.1297 \cdot educyrs. \quad (2)$$

B. Other control variables

The inclusion of our controls can help with omitted variable bias. First off, age serves as a proxy for labor market experience, which tends to increase with time and often correlates positively with both educational attainment and income. Failing to control for age could lead to an inflated estimate of the education coefficient, as older individuals might command higher wages due to experience rather than schooling per se. Secondly, gender is included to capture sex-based wage disparities that, if unaccounted for, could confound the relationship between education and income. Similarly, the NonWhite variable helps account for racial differences in access to educational and occupational opportunities, which may systematically affect wage outcomes. Finally, uhrswork reflects the intensity of labor supply. Because wages are partially determined by the number of hours worked, controlling for this variable helps isolate the effect of education from variation in work effort. Collectively, these controls improve model specification and help ensure that the estimated return to education is not driven by omitted demographic or behavioral factors.

Appendix. Table 6 presents the pairwise correlation coefficients among the variables included in the wage regression model: log wages (lnwage), years of education (educyrs), age, race (NonWhite), gender (female), and usual hours worked per week (uhrswork). The matrix indicates that there is no evidence of perfect collinearity among any pair of variables. Even though we have a coefficient of 0.586 between lnwage and uhrswork, indicating that hours worked is a major determinant of income. However, in our VIF result (Appendix Table 7) all the control variables show VIF less than 5, which is an indication of non-multicollinearity in the model.

C. Multivariate Regression Model with Interaction Term

Below is the multivariate regression function we got from Appendix Table 8:

$$\log(wage) = \beta 0 + \beta 1 \; educyrs + \beta 2 \; age + \beta 3 \; female + \beta 4 \; NonWhite + \beta 5 \; uhrswork + \beta 6 \; femaleuhrswork + u, \quad (3)$$

$$\log(wage)=6.371+0.097\cdot educyrs+0.015\cdot age-0.489\cdot female+0.057\cdot NonWhite+0.049\cdot uhrswork+0.009\cdot femaleuhrswork, \quad (4)$$

In the empirical model presented, the interaction term femaleuhrswork captures how the effect of usual hours worked per week on log wages differs by gender. The estimated coefficient on this interaction is 0.0089, and it is statistically significant at the 1% level, indicating a robust relationship. This positive coefficient suggests that the marginal return to an additional hour worked per week is slightly higher for women compared to men. Specifically, while the base effect of hours worked (uhrswork) on log wages is 0.0494—indicating that each additional hour worked is associated with approximately a 4.94% increase in wages for men—the total effect for women is the sum of the base effect plus the interaction term, i.e.,

$$\text{Total effect for women}=0.0494+0.0089=0.0583, \quad (5)$$

Thus, for women, each additional hour worked is associated with a 5.83% increase in wages, holding other variables constant. From a substantive perspective, this finding implies that women may experience slightly higher proportional wage gains from increasing their labor supply compared to men. This could reflect a number of underlying dynamics—such as differences in occupation sorting, hours flexibility premiums, or labor market discrimination—though such mechanisms would require further investigation. Including this interaction term helps account for gender-based heterogeneity in labor market returns and ensures that the estimated effects of both gender and hours worked are not conflated. Its inclusion therefore enhances the model's ability to isolate the true effect of education on wages by better controlling for variation in labor supply across demographic groups.

For control variables other than the interactive term, the coefficient on age is 0.0154, indicating that, on average, each additional year of age is associated with a 1.54% increase in log wages, holding other factors constant. The female dummy variable is associated with a large negative coefficient of −0.4892, implying that women earn approximately 48.92% less than men, ceteris paribus. This pronounced gender gap in earnings points to systemic disparities that persist despite controlling for key human capital factors. Turning to NonWhite, the coefficient of 0.0566 suggests that non-white workers earn roughly 5.66% more than their white counterparts, while this finding may seem counterintuitive, it may reflect unobserved factors such as regional labor market dynamics or industry-specific concentrations among racial groups. The coefficient on uhrswork is 0.0494, indicating that each additional hour worked per week is associated with a 4.94% increase in wages, which aligns with expectations that more labor input typically translates into higher earnings, especially if additional hours reflect overtime or productivity-based compensation. Finally, the intercept term is estimated at 6.3715, representing the predicted log wage for a baseline individual (male, white, with zero education and hours worked), and this model overall explains approximately 44% of the variation in log wages (R-squared = 0.4398), suggesting a reasonably good fit in capturing wage determinants across the sample.

V. Instrumental Variable Regression

A. Endogeneity of Education

In examining the causal impact of education on wages, a major methodological challenge arises from the potential endogeneity of educational attainment. Years of education (educyrs) may be correlated with unobserved determinants of wages, such as innate ability, family background, or personal motivation. To address this concern, I employ an instrumental variables (IV) approach and use quarter of birth (birthqtr) as an instrument for educational attainment. The economic rationale for this choice draws from institutional features of the education system: in many jurisdictions, school entry laws tie the age of enrollment to calendar cut-off dates. Consequently, individuals born earlier in the year are likely to start school at a younger age, affecting the total number of years they remain in school before reaching the legal dropout age. Therefore, in this paper, I generate a set of three binary variables (q2, q3, q4) representing the second, third, and fourth quarters of birth, eaving the first quarter as the reference category. The F-test(F=18.59) shown in Appendix. Table 9 also proves that instruments q2, q3 and q4 are scientifically significant to be used.

B. Tests for Validity and Endogeneity of Instrumental Variable

To formally test for endogeneity, we employed the Durbin-Wu-Hausman (DWH) test. The null hypothesis of this test is that the suspect regressor—educyrs—is exogenous (i.e., uncorrelated with the error term). The test returned a chi-squared statistic of 6.8615 and a p-value of 0.0088, leading us to reject the null hypothesis at the 1% significance level. This result which is provided in Appendix. Table 10 reveals strong evidence that educyrs is endogenous, thus validating the need for instrumental variable (IV) estimation.

Furthermore, to check whether our instruments (q2, q3, q4) are theoretically valid, we use the first-stage regression, these instruments yield an F-statistic of 19.02, surpassing the commonly accepted threshold of 10, indicating that the instruments are strong and good to use.

Last but not least, is the test for overidentifying restrictions shown in Appendix. Table 12—used to evaluate whether the instruments are uncorrelated with the structural error term—which yields p-values of 0.1019. Since we fail to reject the null hypothesis that the instruments are valid, this provides additional support for the appropriateness of the instruments in the IV specification.

C. IV Regression Results

Table 2. Instrumental Variable (2SLS) Regression Results for Log Wages

lnwage	Coef.	Robust SE	z	P-val	CI_Lower	CI_Upper
educyrs	0.208	0.045	4.655	0.000	0.120	0.295
age	0.014	0.000	30.358	0.000	0.013	0.015
female	-0.479	0.009	-54.117	0.000	-0.496	-0.462
NonWhite	0.014	0.017	0.818	0.413	-0.020	0.048
uhrswork	0.047	0.001	52.088	0.000	0.045	0.049
femaleuhrswork	0.007	0.001	10.705	0.000	0.006	0.009
Constant	5.008	0.551	9.087	0.000	3.928	6.088

$$\log(wage) = 5.01 + 0.208\,educyrs + 0.014\,age - 0.479\,female + 0.014\,NonWhite + 0.047 \cdot uhrswork + 0.007(female\text{-}uhrswork), \tag{6}$$

Standard errors: (0.551) (0.045) (0.000) (0.009) (0.017) (0.001) (0.001)

The coefficient on educyrs, which represents the causal effect of education on log wages (lnwage), is 0.208, statistically significant at the 1% level. This suggests that, holding other factors constant, each additional year of schooling increases wages by approximately 20.8%, which is notably larger than the OLS estimate. This inflation is consistent with attenuation bias in the OLS estimate caused by endogeneity of education. The coefficient on age is 0.0143, implying that each additional year of age increases wages by about 1.43%, possibly reflecting accumulated labor market experience. The gender dummy female shows a significant negative coefficient of -0.479, indicating that women earn about 47.9% less than men, ceteris paribus. The variable uhrswork is positively associated with wages, with a coefficient of 0.047, implying a 4.7% wage increase per additional hour worked per week. The interaction term femaleuhrswork is also positive and statistically significant (coefficient: 0.0073), suggesting that the marginal effect of an additional hour of work on wages is 0.73 percentage points higher for women than for men. This indicates that although women have a lower average wage level, their wage returns to increased labor supply may be stronger at the margin.

Notably, the coefficient on NonWhite is statistically insignificant in the IV model, suggesting that racial differences in wages may not be robust once education and other controls are properly instrumented. Finally, the constant term of 5.008 can be interpreted as the predicted log wage for a baseline individual, though it has limited practical interpretation on its own.

D. Summary and Model Comparison

Table below presents the results from all our regression models generated in this paper. The coefficient on educyrs in the IV regression is 0.208, which implies that each additional year of education is associated with a 20.8% increase in log wages, holding other variables constant. This effect is statistically significant at the 1% level, as indicated by a robust standard error of 0.045 and a p-value of 0.000.

This estimated return to education is notably larger than the coefficient from the multiple OLS regression model (0.097), suggesting that OLS underestimates the true causal impact of education due to endogeneity—possibly arising from omitted ability bias or measurement error in schooling. The IV model addresses this by leveraging exogenous variation in education induced by birth quarter, following the approach of Angrist and Krueger (1991). Their findings also pointed to a positive relationship between quarter of birth, schooling attainment, and wages, though they acknowledged that the quarter-of-birth instrument may explain only a limited portion of the variation in education.

Our result aligns closely with the pattern reported in Angrist and Krueger (1991), who observed statistically significant differences in wages linked to birth quarter and schooling, even if the magnitude was modest. Furthermore, the larger IV estimate supports the hypothesis that measurement error and omitted variables in the OLS model likely biased the coefficient downward, a concern also raised by Sorel and Shinners (2019). They reported a decline in the education coefficient from 12.6% in a simple model to 5.7% in a controlled model, underscoring the influence of confounding variables such as gender and race.

Variable	SimpleL~r	SimpleN~r	Multiple	IVreg
educyrs	6984.065	0.125	0.097	0.208
	16.876	0.000	0.000	0.045
	0.000	0.000	0.000	0.000
age			0.015	0.014
			0.000	0.000
			0.000	0.000
female			-0.489	-0.479
			0.008	0.009
			0.000	0.000
NonWhite			0.057	0.014
			0.002	0.017
			0.000	0.413
uhrswork			0.049	0.047
			0.000	0.001
			0.000	0.000
femaleuhrswork			0.009	0.007
			0.000	0.001
			0.000	0.000

Constant	-45781.648	8.526	6.371	5.008
	242.976	0.005	0.007	0.551
	0.000	0.000	0.000	0.000
N	1,548,402	1,548,402	1,548,402	1,548,402
rmse	62341.984	1.194	0.936	0.991
r2	0.100	0.088	0.440	0.373
r2_a	0.100	0.088	0.440	0.373
F	1.71e+05	1.50e+05	1.13e+05	89665.223

Table 3. Regression Models Comparison

VI. Further Research

To improve and extend this research in the future, several directions could be pursued. While the current analysis uses quarter of birth as an instrument for education, future work could explore alternative or additional instruments—such as changes in compulsory schooling laws or geographic variation in school availability—to strengthen identification and address concerns of weak instrument bias. Moreover, expanding the model to include longitudinal data would allow for fixed effects estimation, controlling for time-invariant unobserved heterogeneity at the individual level, thus improving causal inference. Lastly, linking educational attainment to non-wage outcomes—such as health, employment stability, or job satisfaction—would provide a broader picture of the value of education and yield richer policy insights beyond earnings alone.

References

[1] Angrist, J. D., & Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics, 106*(4), 979–1014. https://doi.org/10.2307/2937954

[2] Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago Press.

[3] Biewen, M., & Fitzenberger, B. (2005). Covariance structure of male wages in West Germany, 1975 – 1995: A semiparametric analysis of changes in distribution, inequality and mobility. *Review of Economic Studies, 72*(2), 665–697. https://doi.org/10.1111/0034-6527.00344

[4] Card, D. (1999). The causal effect of education on earnings. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (Vol. 3, pp. 1801–1863). Elsevier. https://doi.org/10.1016/S1573-4463(99)03011-4

[5] Heckman, J. J., Lochner, L. J., & Todd, P. E. (2006). Earnings functions, rates of return and treatment effects: The Mincer equation and beyond. In E. A. Hanushek & F. Welch (Eds.), *Handbook of the economics of education* (Vol. 1, pp. 307–458). Elsevier. https://doi.org/10.1016/S1574-0692(06)01007-5

[6] Mincer, J. (1974). *Schooling, experience, and earnings*. New York, NY: National Bureau of Economic Research.

[7] Oreopoulos, P. (2006). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review, 96*(1), 152 – 175. https://doi.org/10.1257/000282806776157641

[8] Psacharopoulos, G., & Patrinos, H. A. (2018). Returns to investment in education: A decennial review of the global literature. *Education Economics, 26*(5), 445–458. https://doi.org/10.1080/09645292.2018.1484426

[9] Sorel, T., & Shinners, B. (2019). The relationship between education and wages in Georgia. *Modern Perspectives in Economics and Business, 1*(1), 21–34.

Appendix

Table 4. Simple Regression Model

Variable	Coef.	Robust SE	t	P-value	95% CI Lower	95% CI Upper
educyrs	6984.065	23.738	294.22	0.000	6937.54	7030.59
cons	-45781.65	312.673	-146.42	0.000	-46394.48	-45168.82

Table 5. Nonlinear Simple Regression

Variable	Coef.	Robust SE	t	P-value	95% CI Lower	95% CI Upper
educyrs	0.1252	0.0004	342.95	0.000	0.1245	0.1259
cons	8.5256	0.0053	1605.88	0.000	8.5152	8.5360

Table 6. Pairwise Correlation Matrix Between Variables

	lnwage	educyrs	age	NonWhite	female	uhrswork
lnwage	1.0000					
educyrs	0.2972*	1.0000				
	0.0000					
age	0.2328*	0.0628*	1.0000			
	0.0000	0.0000				
NonWhite	0.0577*	0.0558*	0.0666*	1.0000		
	0.0000	0.0000	0.0000			
female	-0.1545*	0.0547*	-0.0026*	-0.0228*	1.0000	
	0.0000	0.0000	0.0014	0.0000		
uhrswork	0.5857*	0.1029*	0.0585*	0.0267*	-0.1991*	1.0000
	0.0000	0.0000	0.0000	0.0000	0.0000	

Table 7. Variance Inflation Factor (VIF) Diagnostics for Control Variables

Variable	VIF	1/VIF
uhrswork	1.06	0.9446
female	1.05	0.9542
educyrs	1.02	0.9778
age	1.01	0.9894
NonWhite	1.01	0.9921
Mean VIF	1.03	

Table 8. Multivariate Regression Model

	Coef.	Robust SE	t	P-val	CI Lower	CI Upper
educyrs	0.097	0.000	326.65	0.000	0.10	0.10
age	0.015	0.000	271.64	0.000	0.02	0.02
female	-0.489	0.008	-63.88	0.000	-0.50	-0.47
NonWhite	0.057	0.002	29.21	0.000	0.05	0.06
uhrswork	0.049	0.000	366.81	0.000	0.05	0.05
femaleuhrswork	0.009	0.000	46.26	0.000	0.01	0.01
Constant	6.371	0.007	895.77	0.000	6.36	6.39

Table 9. First-Stage Regression of Education on Quarter-of-Birth Dummies

Variable	Coef.	Std. Err.	t	P-value	95% CI Lower	95% CI Upper
q2	0.0595	0.0082	7.25	0.000	0.0435	0.0755
q3	0.0194	0.0086	2.25	0.025	0.0026	0.0363
q4	-0.1435	0.0068	-21.03	0.000	-0.1568	-0.1302
cons	14.0677	0.0048	2910.86	0.000	14.0582	4.0772

Table 10. the Durbin-Wu-Hausman (DWH) test for Endogeneity

Test Type	Statistic	P-value
Robust score χ2(1)	6.8615	0.0088
Robust regression F(1, 1548394)	6.8615	0.0088

Table 11. First-Stage F Statistic and R-Squared

Variable	R2	Adj. R2	Partial R2	F-statistic	P-value
educyrs	0.0231	0.0231	0.0000	19.0219	0.0000

Table.12 Overidentification Test

Test Type	Chi-square (df=2)	P-value
Score test	4.5682	0.1019

Table 13. First-Stage Regression

Variable	Coef.	Robust SE	t	P-value	95% CI Lower	95% CI Upper
age	0.0105	0.00015	68.66	0.000	0.0104	0.0107
female	-0.9276	0.0154	-59.77	0.000	-0.9572	-0.8981
NonWhite	-0.3212	0.0067	-47.98	0.000	-0.3344	-0.3081
uhrswork	0.0584	0.0008	73.88	0.000	0.0568	0.0600
femaleuhrswork	0.0065	0.0002	30.45	0.000	0.0061	0.0069
q2	0.0595	0.0082	7.24	0.000	0.0435	0.0755
q3	0.0194	0.0086	2.25	0.025	0.0026	0.0363
q4	-0.1435	0.0068	-21.03	0.000	-0.1568	-0.1302
cons	12.3326	0.0136	905.22	0.000	12.2960	12.3499