WEP Warwick Evans Publishing

International Journal of Global Economics and Management

ISSN: 3005-9690 (Print), ISSN: 3005-8090 (Online) | Volume 8, Number 3, Year 2025
DOI: https://doi.org/10.62051/ijgem.v8n3.09
Journal homepage: https://ijgem.org

Effects of Working Hours Per Week on Wages

Xinzi Li

School of Mathematics and Physics, Xi'an Jiaotong-Liverpool University, Suzhou, 215123, China
Xinzi.Li22@student.xjtlu.edu.cn

ABSTRACT

This study examines the nonlinear relationship between working hours and wages using cross-sectional data from the American Community Survey (ACS), while accounting for variables such as age, gender, race, and language proficiency. By constructing three progressively expanded regression models, the findings reveal a significant positive correlation between working hours and wages, albeit with diminishing marginal returns. Additionally, age exhibits a concave relationship with wages, where earnings initially rise with age before plateauing. Educational attainment further moderates this relationship, with highly educated individuals benefiting more from additional work hours but also experiencing greater productivity losses due to overwork. The study also highlights persistent wage disparities based on gender and race, with female and non-white workers facing significant income gaps. The inclusion of quadratic terms improves model fit, providing empirical insights for labor market policy formulation.

KEYWORDS

Working hours; Wage determination; Nonlinear regression; Educational attainment; American Community Survey (ACS)

1. MOTIVATION AND RESEARCH QUESTION

The causation between personal work hours and wages is a crucial issue that has been extensively studied in labor economics. Some investigations indicated that there is a positive correlation between work hours and income. According to Daniel, hourly wages for workers in an extended period overtop those in the regular time by 32% [1]. Nevertheless, other experimental researches suggest that working hours affect salaries negatively. Jieunreported that hourly wages would be lifted by 3.38% when the work time is reduced by nearly 51 minutes a week [2]. In 2022, the recovery in average work hours fostered GDP to sustainably grow by approximately 2.6%. In contemporary society, the significance of studying the relationship between working hours and wages is to reveal the advantages and disadvantages of working hours to individual income, while also demonstrating that work hours are important driving factors of productivity and economic development at the macro level. On the basis of the existing evidence, it is crucial to analyze the influence of individual working hours on wages and to conduct further research.

In accordance with the scholarly investigation by Bick, Blandin, and Rogerson, it has been ascertained that there exists a non-monotonic relationship between hourly wages and the typical weekly working hours [3]. Concurrently, the research conducted by Ozhamaratli, Kitov, and Barucca delineates a nonlinear relationship between age and income, characterized by a rapid increase in income during early career stages, followed by a period of stability around the aver- age income level, and ultimately a decline converging towards the average retirement income [4]. To bolster the rigor of our analytical framework, we include quadratic terms for both age and working hours in the analysis to capture potential non-linear effects.

Additionally, building on cross-sectional data from the American Community Survey, this research examines the complex relationship between working hours and wages, incorporating key variables such as age, gender, race, and language proficiency. Research by Fern´andez-Val et al. in arXiv analyzes the impact of working hours annually on the distribution of real annual earnings in the U.S. from 1976 to 2019, highlighting significant differences in how female employees behave and their effects on income inequality [5]. A study in Labour Economics by Goos and Manning explores how population and skill-based differences contribute to wage gaps, providing essential context for understanding inequalities across groups [6]. On the account of various careers leads non-linear relationships between working hours and earnings, the population is divided into different educational levels, in order to investigate how different jobs connects corresponding working hours, draws out how wages are affected. Special attention is given to population inequalities, especially those arising from gender and racial differences, providing insights for policy making.

2. DATA DESCRIPTION

2.1. Data Sources

The American Community Survey Public Use Microdata Sample (PUMS) was used to gather information on individuals' wages, usual hours worked per week past 12 month, age, race, gender, language other than English spoken, and educational attainment in the US in 2023. This website is an excellent place to find high-quality statistics about the country's population and economy, which is useful for figuring out how individual wages relate to other factors.

From the whole dataset, we chose a one-year sample for our study. For data selection, a random sample of 10,000 observations was initially drawn from the entire dataset. After cleaning the data by removing any invalid or incomplete rows, a total of 9,514 valid observations were retained for analysis. This cleaning process ensured that the final dataset was both representative and reliable for exploring the relationship between working hours and wages.

2.2. Variable Measurement

Table 1. The definition of variables

No.	Variable	Definition
1	$\ln_{wage}$	The natural logarithm of an individual's wages in the past 12 months.
2	WorkHour	The average number of hours worked per week.
3	Age	The age of targeted people in New York
4	LOTE	Language Other Than English
5	Race	0-all other races, 1-white only, dummy variable
6	Gender	0-male, 1-female, dummy variable

Table 1 defines the variables used in the analysis, which include the natural log of wages (lnwage), weekly work hours (WorkHour), Age, whether a language other than English is spoken (LOTE), and dummy variables for Race (white only) and Gender (female).

2.2.1. Wage ($\ln_{wage}$)

The dependent variable in the regression is the natural logarithm of wage ($\ln_{wage}$ wage). From figure 1, we can observe that the distribution of wage is highly right- skewed. Most wages are concentrated in the lower range, while a small number of high wages create a long tail. This kind of skewness can cause issues in regression analysis, as it may violate the assumption of normally distributed residuals, affecting the robustness and accuracy of the model. By taking the natural logarithm of wages ($\ln_{wage}$), as shown in the figure 2, the distribution becomes more symmetric and closer to a normal distribution, which aligns better with the assumptions of many statistical models.

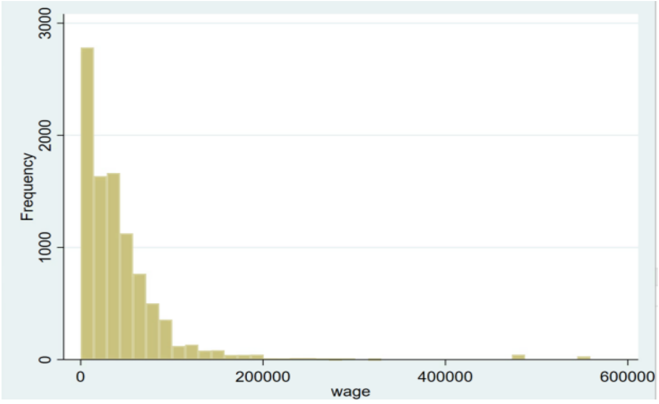

Figure 1. Wage Distribution

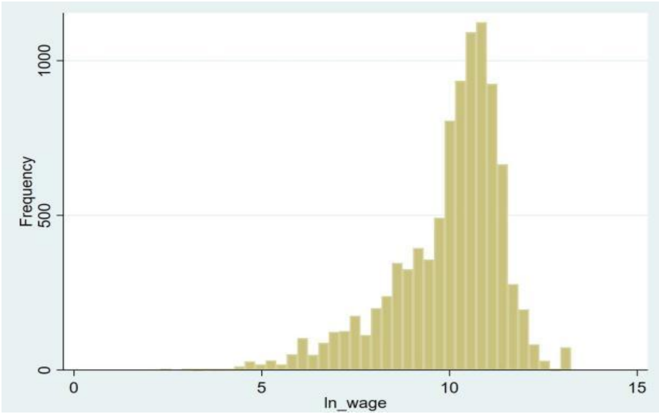

Figure 2. lnwage Distribution

2.2.2. Work hour (WorkHour)

In the analysis, work duration is identified as a key independent variable significantly influencing hourly wages. Bick, A., Blandin, A., and Rogerson explored the intricate relationship between work hours and productivity, finding that reducing work hours can increase hourly wages or productivity [7]. Their study indicates that hourly wages initially rise with work hours up to 50 per week, after which they decline due to fatigue and diminishing marginal productivity returns from extended work. This highlights the non-linear interplay between work duration and wage income.

2.2.3. Age

The explanatory variable of age accounts for the disparities in salary income associated with age. In their study, Ozhamaratli, Kitov, and Barucca conducted a thorough analysis of the relationship between age and income distribution in both the United Kingdom and the United States [4]. The study demonstrates that during the initial years of an individual's professional life, there is a swift increase in income. This is followed by a period of stabilization around the average income level, which persists throughout the remainder of the individual's career until retirement. The study highlights the complex, non-linear relationship between work duration and wage income.

2.2.4. Language other than English (LOTE)

The explanatory variable *LOTE* highlights wage differences between individuals who speak only English and those who speak both English and other languages. In countries where English is the primary language, bilingualism or multilingualism can provide advantages in special sectors, such as international business or translation services. Fry and Lowell observed that those who speak other languages than English can lead to more wages in particular careers, it is dependent on the needs for certain language abilities [8]. Similarly, Saiz and Zoido found that those who graduate from colleges in the U.S. that speak a second language earn around 2-3% more than average [9]. Therefore, *LOTE* is included as an explanatory variable in this analysis to examine how it impacts wage disparities across industries and demographic groups.

2.2.5. Race

The variable *Race* is introduced as a dummy variable to differentiate between White individuals (coded as 1) and non-White individuals (coded as 0). Data from the U.S. Department of Labor indicates that, on average, Black workers earn around three fourths of what White workers earn, highlighting a significant wage disparity, therefore it shows persistent racial wage gaps in the labor market [10]. Additionally, Kochhar and Cilluffo found that the disparity of earnings among races in 2016 are greater compared to 1970, even though the income differences in blacks and whites become marginally closer [11]. Hence, we expect wages to be positively associated with being White.

2.2.6. Gender

The independent variable, *Gender*, is introduced as a dummy variable to distinguish between female and male workers, with the value 1 assigned to females. This variable allows us to explore gender differences in the relationship between men and women. Amalia reported that women have low representation in high-pay, high-status, elite careers on account of required long work hours [12]. As a result, it is predicted that Gender would negatively moderate wages.

2.3. Descriptive Statistics of the Variables

The descriptive statistics of variables are presented in Table 2.

Table 2. Description of Variables

Variable	Mean	Std. dev.	Min	Max
$\ln_{wage}$	9.97	1.48	2.30	13.23
WorkHour	36.42	13.88	1	99
Age	40.65	16.37	16	94
LOTE	1.80	0.40	1	2
Gender	1.44	0.50	1	2
Race	2.76	2.91	1	9

3. EMPIRICAL MODEL

3.1. Model 1

This baseline model examines the direct effects of working hours, age, language proficiency, race, and gender on wages. Thus, we build our basic model:

$$\ln_{wage} = \beta_0 + \beta_1 WorkHour + \beta_2 Age + \beta_3 LOTE + \beta_4 Race + \beta_5 Gender + u_i, \quad \text{for } i = 1 \ldots n \tag{1}$$

3.2. Model 2

The inclusion of Age^2 captures the potential non-linear relationship between age and wages, where wages might increase with age but at a decreasing rate.

$$lnwage = \beta 0 + \beta 1 WorkHour + \beta 2 Age + \beta 3 Age2 + \beta 4 LOTE$$
$$+ \beta 6 Race + \beta 7 Gender + ui, fori = 1 ... n \quad (2)$$

3.3. Model 3

The addition of *WorkHour*^{2} accounts for the diminishing returns to wages with in- creasing work hours, as fatigue and productivity loss may reduce wage growth beyond certain thresholds.

$$lnwage = \beta 0 + \beta 1 WorkHour + \beta 2 WorkHour2 + \beta 3 Age + \beta 4 Age2 + \beta 5 LOTE$$
$$+ \beta 6 Race + \beta 7 Gender + ui, fori = 1 ... n \quad (3)$$

4. PRESENTATION OF RESULTS AND INFERENCE

Table 3. Coefficients and Standard Errors for the Models 1, 2, and 3

Regressor	Model Coefficient	1 Std	Model Coefficient	2 Std	Model Coefficient	3 Std
WorkHour	0.0540	0.0009	0.0465	0.0009	0.1114	0.0025
WorkHour^2					-0.0008	0.0000
Age	0.0227	0.0007	0.1113	0.0041	0.0899	0.0040
Age^2			0.0010	0.0001	-0.0008	0.0000
LOTE	-0.0914	0.0345	-0.0328	0.0338	-0.0453	0.0325
Race	-0.0443	0.0047	-0.0434	0.0046	-0.0440	0.0044
Gender	-0.0398	0.0241	-0.0613	0.0235	-0.0486	0.0226
Adjusted R^2	0.3935		0.4231		0.4672	
Intercept	7.4205		5.9981		5.3570	

Table 3 presents the estimated coefficients and standard errors for three regression models. The results show that adding quadratic terms for WorkHour and Age in Models 2 and 3 significantly improves the model fit, as evidenced by the increasing Adjusted R^2 values from 0.3935 to 0.4672. Most variables, including WorkHour, Age, LOTE, Race, and Gender, are statistically significant across the models, with their coefficients and significance levels showing some variation as new terms are introduced.

Table 4. Comparison of t values and P > |t| for the Models 1, 2, and 3

Variable	Model 1		Model 2		Model 3	
	t	P > |t|	t	P > |t|	t	P > |t|
WorkHour	61.00	0.000	50.08	0.000	44.94	0.000
WorkHour^2					-15.53	0.000
Age	30.56	0.000	27.35	0.000	22.56	0.000
Age^2			-17.62	0.000	-6.69	0.000
LOTE	-2.65	0.008	-0.97	0.331	-1.40	0.162
Race	-9.36	0.000	-9.39	0.000	-9.92	0.000
Gender	-1.65	0.099	-2.60	0.009	-2.15	0.032

4.1. Model 1

In Model 1, the relationship between working hours, age, language proficiency, race, and gender on wages ($\ln_{wage}$) is explored. The null and alternative hypotheses are as follows:

$$H0{:}\,\beta 1 = 0 against H1{:}\,\beta 10 \quad (4)$$

The coefficient for *WorkHour* is 0.0540, with a t-value of 61.00 and a p-value of 0.000. This indicates a highly significant positive relationship between working hours and wages. For each additional hour worked, wages increase by approximately 5.4%. The p-value is 0.000, which is less than 0.05, so we reject the null hypothesis and conclude that working hours significantly affect wages.

$$H0{:}\,\beta j = 0 against H1{:}\,\beta j 0, j = 2{,}3{,}4{,}5 \quad (5)$$

The analysis shows that *Age*, *LOTE*, and *Race* have high t-values and low p-values, leading to the rejection of the null hypothesis at the 5% significance level. The coefficient for *Age* (0.0227) indicates a positive relationship between age and wages, with each additional year of age increasing wages by 2.2%, reflecting the impact of experience and seniority. The results reveal a significant negative relationship between proficiency in languages other than English and wages, with limited proficiency leading to lower earnings. Additionally, a significant racial wage gap exists, where non-white individuals earn less than white counterparts. Regarding gender, the coefficient indicates a marginally significant negative effect on wages at the 10% significance level. However, with a p-value greater than 0.05, gender does not have a statistically significant effect on wages at the 5% significance level in Model 1.

4.2. Model 2

Model 2 builds upon the findings from Model 1 by introducing a nonlinear term, Age^2, to capture the complex relationship between age and wages as suggested by Ozhamaratli, Kitov, and Barucca [3]. This adjustment acknowledges that the wage increase with age may not be uniform but rather decelerates over time, which is supported by the scatter plot in Figure 3 and Figure 4.

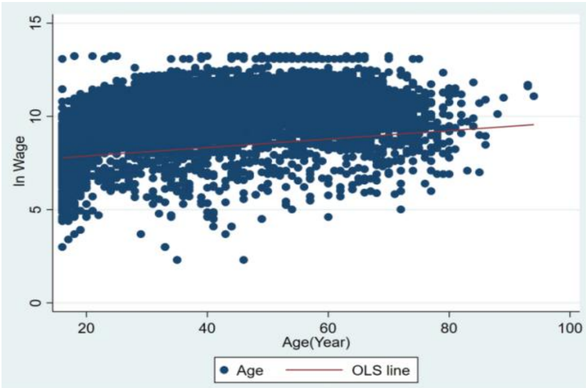

Figure 3. Scatterplot of $\ln_{wage}$ and age with a log OLS regression line

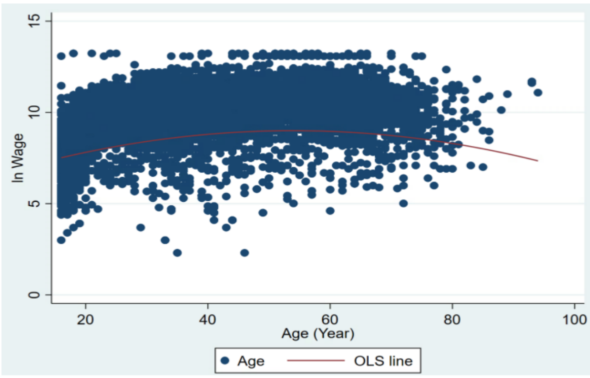

Figure 4. Scatterplot of $\ln_{wage}$ and age, age^2 with a log OLS regression line

In Model 2, the Age^2 term is included to account for this nonlinearity, and the statistical tests for its significance are set as follows:

$$H0{:}\beta 3 = 0 against H1{:}\beta 30 \quad (6)$$

The addition of Age^2 to Model 2 is significant, with a coefficient of 0.0010, a t-value of -17.62, and a p-value near zero, establishing a quadratic relationship. The linear age term's coefficient also rises to 0.1113, confirming a concave age-wage relationship where wage growth with age initially increases but then decelerates.

WorkHour's coefficient modestly decreases from Model 1 to Model 2, influenced by the age-related term. The racial wage gap persists, as indicated by a coefficient for Race of -0.0434. Gender's impact on wages becomes significant in Model 2, with a coefficient of -0.0613, a t-value of -2.61, and a p-value of 0.009, suggesting a more pronounced gender wage gap than in Model 1.

The coefficient for *LOTE* is -0.0328, with a t-value of -0.97 and a p-value of 0.331, indicating that controlling for age lessens the negative wage effect of *LOTE*. This implies that age partially accounts for the wage impact of non-English language proficiency.

The inclusion of Age^2 improves the model fit, as evidenced by an increase in the R-square value from 0.3935 to 0.4231, enhancing the model's explanatory power compared to the initial model.

4.3. Model 3

Model 3 delves into the relationship between working hours, age, language proficiency, race, and gender on wages, considering the nonlinearity of working hours as suggested by Bick, Blandin, and Rogerson [3]. This research indicates that the effect of additional work hours on wages may diminish as work hours increase due to factors like fatigue or productivity constraints, which is supported by the scatter plot in Figure 5 and Figure 6.

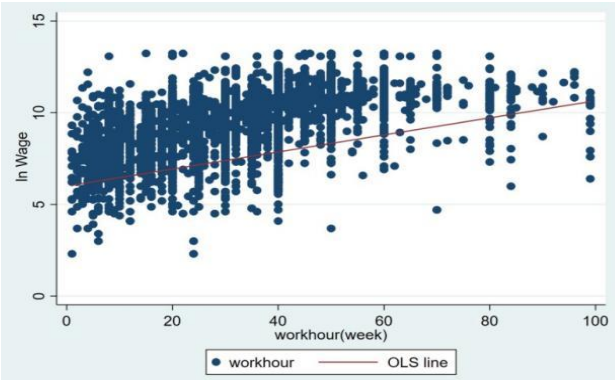

Figure 5. Scatterplot of $\ln_{wage}$ and WorkHour with a log OLS regression line

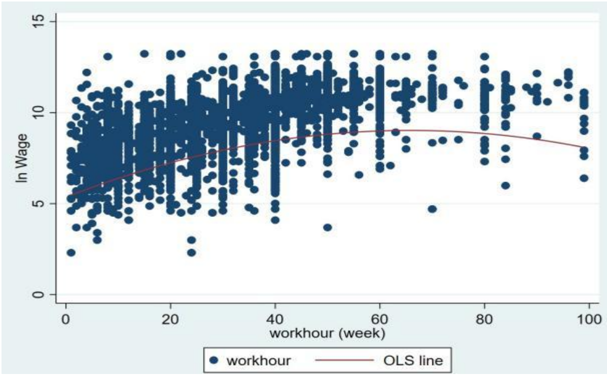

Figure 6. Scatterplot of $\ln_{wage}$ and WorkHour, *WorkHour²* with a log OLS regression line

To test the significance of the nonlinear term for working hours, *WorkHour²*, we set the hypotheses as:

$$H0\colon \beta 2 = 0 against H1\colon \beta 20 \tag{7}$$

The coefficient for *WorkHour* is 0.1114, with a t-value of 44.56 and a p-value of 0.0000, indicating a significant positive effect of working hours on wages, albeit with diminishing returns as shown by the coefficient for *WorkHour²*. The results for *Age* and *Age²* reaffirm a concave relationship between age and wages, suggesting that wages increase with age but at a decreasing rate.

For *Race* and *Gender*, the outcomes confirm the persistence of racial and gender wage gaps at the 5% significance level. However, the coefficient for *LOTE* is -0.0453, with a t-value of -1.40 and a p-value of 0.1610, indicating that the impact of language proficiency on wages is no longer statistically significant at the 5% level.

In summary, Model 3 confirms that working hours, age, race, and gender significantly determine wages, with the hypothesis tests consistently rejecting the null hypotheses for *WorkHour*, *Age*, *Race*, and *Gender*. This supports the theoretical expectations of diminishing returns to working hours, a lifecycle wage pattern, and persistent wage disparities.

5. EXTENSION

Practically, the influence of work hours on wages is non-linear on account of distinction by career. Daniel claimed that extensive work time leads decline in hourly wages among professional and technical workers, whereas it increases for management personnel [1]. As different occupations require different levels of education, we keep other variables in model3 and establish three groups to divide the levels of education from low, moderate to high, which are workers with less than high school degree (Edu≤15, expressed as ‘LE’), high school graduate ($15 < \text{Edu} \leq 19$ years, expressed as ‘ME’) and college and 10 higher graduate ($19 < \text{Edu} \leq 24$ years, expressed as ‘HE’) respectively. From the statistics, approximately 4242 workers in the US have less than a high school degree, while the number of middle education level workers are 2857, and a college degree or more is 3313.

Table 5. Illustrations for Edu

Value of Edu	Definition	Value of Edu	Definition
01	No schooling completed	13	Grade 10
02	Nursery school, preschool	14	Grade 11
03	Kindergarten	15	12th grade - no diploma
04	Grade 1	16	Regular high school diploma
05	Grade 2	17	GED or alternative creden- tial
06	Grade 3	18	Some college, but less than 1 year
07	Grade 4	19	1 or more years of college credit, no degree
08	Grade 5	20	Associate’s degree
09	Grade 6	21	Bachelor’s degree
10	Grade 7	22	Master’s degree
11	Grade 8	23	Professional degree beyond a bachelor’s degree
12	Grade 9	24	Doctorate degree

In order to test the significance of three education levels, the null and alternative hypotheses are as follows:

$$H0\colon \beta 1 - LE = 0 against H1\colon \beta 1 - LE0 \tag{8}$$

$$H0: \beta1 - ME = 0 against H1: \beta1 - ME0 \tag{9}$$

$$H0: \beta1 - HE = 0 against H1: \beta1 - HE0 \tag{10}$$

Table 6. Coefficients and Standard Errors for the Models LE, ME, and HE

	Model LE		Model ME		Model HE	
Regressor	Coefficient	Std	Coefficient	Std	Coefficient	Std
WorkHour	0.1007	0.0038	0.1085	0.0038	0.1104	0.0041
$WorkHour^2$	-0.0007	0.0000	-0.0008	0.0000	-0.0009	0.0000
Age	0.0638	0.0061	0.0679	0.0062	0.1022	0.0070
Age^2	-0.0005	0.0001	-0.0006	0.0001	-0.0010	0.0001
LOTE	-0.1294	0.0504	0.0623	0.0550	-0.0295	0.0494
Race	-0.0112	0.0066	0.0061	0.0067	-0.0530	0.0101
Gender	-0.0904	0.0378	-0.1536	0.03485	-0.1781	0.0329
Adjusted R^2	0.4473	0.4389		0.4144		
Intercept	5.8391	5.7885		5.7701		

Table 7. Comparison of t values and P > |t| for the Models LE, ME, and HE

Variable	Model LE		Model ME		Model HE	
	t	P > |t|	t	P > |t|	t	P > |t|
WorkHour	26.20	0.000	28.23	0.000	27.18	0.000
$WorkHour^2$	-15.53	0.000	-18.6	0.000	-17.89	0.000
Age	10.32	0.000	10.89	0.000	14.58	0.000
Age^2	-6.69	0.000	-8.53	0.000	-12.69	0.550
LOTE	-2.57	0.010	1.13	0.257	-0.60	0.000
Race	-1.70	0.089	0.99	0.324	-5.28	0.000
Gender	-2.39	0.017	-4.41	0.000	-55.41	0.000

The coefficient for *WorkHour* is positive for all groups, suggesting that higher wages are typically associated with longer working hours. With a coefficient of 0.1104, the highly educated group (HE) has the highest effect, while the ME and LE groups have coefficients of 0.1085 and 0.1007, respectively. This implies that the biggest earnings benefits from longer work hours are experienced by those with more education. This pattern likely reflects the higher marginal productivity of educated workers, as their additional hours contribute more significantly to tasks requiring advanced skills or specialized knowledge.

The non-linear link between working hours and salaries is highlighted by the inclusion of the *WorkHour*^{2} which has negative coefficients for all groups. As working hours grow, this indicates decreasing returns to salaries, especially in the HE group where the coefficient is -0.0009. The result implies that although highly educated people benefit more from working extra hours at first, their returns also decrease the most as work hours increase. These declining returns might be brought on by weariness, decreased output, or the higher potential cost of time for highly qualified workers.

In summary, the study demonstrates that although working longer hours can increase wages for workers of all educational backgrounds, the extent of this benefit and the rate at which returns decline differ significantly. Employees with higher levels of education gain sufficiently from more hours worked, but they are the easiest to suffer from overwork. This emphasizes how crucial it is to strike a balance between productivity and burden, especially for people in high-skilled professions.

6. CONCLUSION AND DISCUSSION

This study examines how working hours affect pay while taking factors such as age, LOTE, race, and gender into account. In order to identify correlations, we created a non-linear regression model using cross-sectional data from the American Community Survey. Our results show that earnings are positively correlated with age, working hours, and educational attainment, with higher returns shown in highly educated groups. Educational attainment is included as it influences wages by reflecting skills and careers, reduces demographic disparities like race and gender gaps, and provides insights into systemic biases in the labour market. The model's fit was enhanced by adding quadratic factors for age and working hours, which supported the concave relationship between age and wages as well as the diminishing rewards for longer work hours.

However, some limitations exist in the non-linear regression model. For instance, gender appears to have a more significant influence on wage disparities than initially anticipated, while the variable for race becomes statistically insignificant in the high-education group. Additionally, the current model does not fully account for variables such as work experience, which could better explain wage variations. These aspects should be taken into account in future studies to improve the analysis's robustness.

To summarize, while this study emphasizes the role of working hours in determining wages, the broader implication lies in addressing wage inequalities. Policymakers should concentrate on expanding educational opportunities, reducing gender-based and race-based gaps, and promoting equitable wage policies to achieve economic development. Further exploration into industry-specific wage determinants will also refine our understanding of this complex relationship.

REFERENCES

[1] Daniel H. (1998) How hours of work affect occupation earning?. https://www.bls.gov/opub/mlr/1998/10/art2full.pdf

[2] Jieun L. (2022) Effects of a national work hours restriction in a high hours country. https://doi.org/10.2478/izajolp-2022-0006

[3] Bick, A., Blandin, A. and Rogerson, R. (2020) 'Hours and Wages', Working Paper No. 26722, National Bureau of Economic Research, Cambridge, MA. http://www.nber.org/papers/w26722

[4] Ozhamaratli, F., Kitov, O. and Barucca, P.(2022). A generative model for age and income distribution. EPJ Data Science, 11, 4.

[5] Fernández-Val, I., Liu, Y., Van Vuuren, A. and Vella, F. (2020). The impact of hours worked on annual earnings inequality. https://arxiv.org/abs/2002.11211

[6] Goos, M. and Manning, A. (2007). Lousy and lovely jobs: The rising polarization of work in Britain. Labour Economics, 14(6), 979–1010.

[7] Bick, A., Blandin, A. and Rogerson, R. (2022). Hours and wages. The Quarterly Journal of Economics, 137(3), 1901-1962.

[8] Fry, R. and Lowell, B.L. (2003). The Value of Bilingualism in the U.S. Labor Market. Industrial and Labor Relations Review, 57(1), 128–140.

[9] Saiz, A. and Zoido, E. (2005). Listening to What the World Says: Bilingualism and Earnings in the United States. The Review of Economics and Statistics, 87(3), 523–538.

[10] U.S. Department of Labor (2023). Earnings: Race and Ethnicity. https://www.dol.gov/agencies/ofccp/about/data/earnings/race-and-ethnicity?utm_source=chatgpt.com

[11] Pew Research Center (2018). Key findings on the rise in income inequality within America's racial and ethnic groups. https://www.pewresearch.org/short-reads/2018/07/12/key-findings-on-the-rise-in-income-inequality-within-americas-racial-and-ethnic-groups/?utm_source=chatgpt.com

[12] Amalia R. M. (2024) Effects of Workplace Competition on Work Time and Gender Inequality. Ilr Review, 77(2), 251–272.

APPENDICES

A.1 STATA code

```
gen age2 = age^2
gen workhour2 = workhour^2
drop if wage == 0
gen LE = cond(edu <= 15, edu, .)
gen HE = cond(edu >= 19, edu, .)
gen ME = cond(inrange(edu, 16, 19), edu, .)
summarize ln_wage age gender LOTE workhour race
regress ln_wage workhour age LOTE race gender
scalar cb1 = _b[age]
scalar cc1 = _b[_cons]
twoway (scatter ln_wage age) (function y=cc1+cb1*x, range(age)), legend(label(1 "Age") label(2 "OLS line")) xtitle("Age(Year)") ytitle("ln Wage")
regress ln_wage workhour age age2 LOTE race gender,vce(robust)
scalar cb2 = _b[workhour]
scalar cc2 = _b[_cons]
twoway (scatter ln_wage workhour) (function y=cc2+cb2*x, range(workhour)), legend(label(1 "workhour") label(2 "OLS line")) xtitle("workhour(week)") ytitle("ln Wage")
regress ln_wage workhour age age2 LOTE race gender,vce(robust)
scalar cb3 = _b[age]
scalar cb4 = _b[age2]
scalar cc3 = _b[_cons]
twoway (scatter ln_wage age) (function y=cc3+cb3*x+cb4*x^2, range(age)), legend(label(1 "Age") label(2 "OLS line")) xtitle("Age(Year)") ytitle("ln Wage")
regress ln_wage workhour workhour2 age age2 LOTE race gender,vce(robust)
scalar cb5 = _b[workhour]
scalar cb6 = _b[workhour2]
scalar cc5 = _b[_cons]
twoway (scatter ln_wage workhour) (function y=cc5+cb5*x+cb6*x^2, range(workhour)), legend(label(1 "workhour") label(2 "OLS line")) xtitle("workhour(week)") ytitle("ln Wage")
regress ln_wage workhour workhour2 age age2 LOTE race gender if LE != . ,vce(robust)
regress ln_wage workhour workhour2 age age2 LOTE race gender if ME != . ,vce(robust)
regress ln_wage workhour workhour2 age age2 LOTE race gender if HE != . ,vce(robust)
```

A.2. Regression output

```
summarize ln_wage age gender LOTE workhour race
```

Variable	Obs	Mean	Std. dev.	Min	Max
ln_wage	9514	9.967549	1.47674	2.302585	13.233
age	9514	40.65409	16.37195	16	94
gender	9514	1.438722	0.4962569	1	2
LOTE	9514	1.801976	0.3985313	1	2
workhour	9514	36.42148	13.87798	1	99
race	9514	2.76172	2.908515	1	6

regress ln_wage workhour age LOTE race gender,vce(robust)

Number of obs = 9,514

F(5, 9508) = 1235.33

Prob > F = 0.0000

R-squared = 0.3938

Adj R-squared = 0.3935

Root MSE = 1.1501

ln_wage	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
workhour	0.0540434	0.000886	61.00	0.000	0.0523066	0.0557802
age	0.0227083	0.0007431	30.56	0.000	0.0212516	0.0241649
LOTE	-0.0914115	0.0345208	-2.65	0.008	-0.1590796	-0.0237434
race	-0.0443475	0.0047376	-9.36	0.000	-0.0536342	-0.0350607
gender	-0.0397825	0.0241212	-1.65	0.099	-0.0870652	0.0075002
_cons	7.420457	0.0931282	79.68	0.000	7.237906	7.603008

regress ln_wage workhour age age2 LOTE race gender,vce(robust)

Number of obs = 9,514

F(6, 9507) = 1163.87

Prob > F = 0.0000

R-squared = 0.4235

Adj R-squared = 0.4231

Root MSE = 1.1216

ln_wage	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
workhour	0.0465113	0.0009288	50.08	0.000	0.0446907	0.0483319
age	0.1112802	0.0040691	27.35	0.000	0.103304	0.1192564
age2	-0.0010316	0.0000466	-22.12	0.000	-0.001123	-0.0009402
LOTE	-0.0328005	0.0337711	-0.97	0.331	-0.0989991	0.0333982

race | -0.0433778 | 0.0046207 | -9.39 | 0.000 | -0.0524353 -0.0343203

gender | -0.0613313 | 0.0235448 | -2.60 | 0.009 | -0.1074841 -0.0151784

_cons | 5.998097 | 0.1112815 | 53.90 | 0.000 | 5.779962 6.216233

regress ln_wage workhour workhour2 age age2 LOTE race gender,vce(robust)

Number of obs = 9,514

F(7, 9506) = 1192.63

Prob > F = 0.0000

R-squared = 0.4676

Adj R-squared = 0.4672

Root MSE = 1.0779

ln_wage | Coefficient | Std. err. | t | P>|t| | [95% conf. interval]

----

workhour | 0.1114318 | 0.0024796 | 44.94 | 0.000 | 0.1065712 0.1162925

workhour2 | -0.0008469 | 0.0000302 | -28.06| 0.000 | -0.000906 -0.0007877

age | 0.0898698 | 0.0039842 | 22.56 | 0.000 | 0.0820598 0.0976798

age2 | -0.0008027 | 0.0000456 | -17.62| 0.000 | -0.000892 -0.0007134

LOTE | -0.0453476 | 0.0324584 | -1.40 | 0.162 | -0.108973 0.0182777

race | -0.0440327 | 0.0044407 | -9.92 | 0.000 | -0.0527374 -0.035328

gender | -0.0486272 | 0.022632 | -2.15 | 0.032 | -0.0929907 -0.0042637

_cons | 5.357001 | 0.1093585 | 48.99 | 0.000 | 5.142635 5.571367

regress ln_wage workhour workhour2 age age2 LOTE race gender if LE != . ,vce(robust)

Number of obs = 4,242

F(7, 4234) = 491.23

Prob > F = 0.0000

R-squared = 0.4482

Adj R-squared = 0.4473

Root MSE = 1.1753

ln_wage | Coefficient | Std. err. | t | P>|t| | [95% conf. interval]

----

workhour | 0.1006907 | 0.0038429 | 26.20 | 0.000 | 0.0931565 0.1082248

workhour2 | -0.0007429 | 0.0000478 | -15.53| 0.000 | -0.0008367 -0.0006491

age | 0.0638458 | 0.0061852 | 10.32 | 0.000 | 0.0517196 0.0759721

age2 | -0.0004893 | 0.0000731 | -6.69 | 0.000 | -0.0006327 -0.000346

LOTE | -0.1294184 | 0.0503911 | -2.57 | 0.010 | -0.2282114 -0.0306254

race | -0.0112485 | 0.0066085 | -1.70 | 0.089 | -0.0242046 0.0017077

gender | -0.0904438 | 0.0378474 | -2.39 | 0.017 | -0.1646446 -0.0162429 _cons | 5.839091 | 0.1678712 | 34.78 | 0.000 | 5.509975 6.168207

regress ln_wage workhour workhour2 age age2 LOTE race gender if ME != . ,vce(robust)

Number of obs = 2,857

F(7, 2849) = 320.17

Prob > F = 0.0000

R-squared = 0.4403

Adj R-squared = 0.4389

Root MSE = 0.90914

ln_wage | Coefficient | Std. err. | t | P>|t| | [95% conf. interval]

----

workhour | 0.108482 | 0.0038426 | 28.23 | 0.000 | 0.1009474 0.1160165

workhour2 | -0.0008344 | 0.0000449 | -18.60| 0.000 | -0.0009224 -0.0007464

age | 0.0679565 | 0.0062416 | 10.89 | 0.000 | 0.0557179 0.0801951

age2 | -0.0005959 | 0.0000699 | -8.53 | 0.000 | -0.0007328 -0.0004589

LOTE | 0.062302 | 0.0550015 | 1.13 | 0.257 | -0.0455448 0.1701488

race | 0.0065697 | 0.0066668 | 0.99 | 0.324 | -0.0065026 0.019642

gender | -0.1535826 | 0.0348482 | -4.41 | 0.000 | -0.2219128 -0.0852524

_cons | 5.788529 | 0.1803118 | 32.10 | 0.000 | 5.434974 6.142084

regress ln_wage workhour workhour2 age age2 LOTE race gender if HE != . ,vce(robust)

Number of obs = 3,313

F(7, 3305) = 335.83

Prob > F = 0.0000

R-squared = 0.4156

Adj R-squared = 0.4144

Root MSE = 0.92347

ln_wage | Coefficient | Std. err. | t | P>|t| | [95% conf. interval]

----

workhour | 0.110405 | 0.0040624 | 27.18 | 0.000 | 0.1024399 0.1183701

workhour2 | -0.000876 | 0.000049 | -17.89| 0.000 | -0.000972 -0.00078

age | 0.1021733 | 0.0070068 | 14.58 | 0.000 | 0.0884353 0.1159114

age2 | -0.0009693 | 0.0000764 | -12.69| 0.000 | -0.0011191 -0.0008195

LOTE | -0.0295149 | 0.0493819 | -0.60 | 0.550 | -0.1263371 0.0673074

race | -0.0530419 | 0.0100553 | -5.28 | 0.000 | -0.0727571 -0.0333267

gender | -0.1781434 | 0.0329398 | -5.41 | 0.000 | -0.2427279 -0.1135588

_cons | 5.770064 | 0.189915 | 30.38 | 0.000 | 5.397701 6.142427