Comparison of Linear and Non-Linear Regression Models in Analyzing Relationship Between Age and Income

Zhanning Sun *

Malvern College, Malvern, United Kingdom

* Corresponding Author Email: Sun_Zhanning@malverncollege.org.uk

Abstract. Currently, many Chinese people are facing the situation of being laid off during their middle age, and the economic situation was bad and it was becoming increasingly difficult to earn money. This made the author want to figure out what factors exactly could affect a person's salary. In particular, the author wants to explore whether there is a certain relationship between age and salary that can be expressed by a function, and attempts to identify other factors that affect salary. In this article, the author investigated the relationship between age and salary and explored the influence of other variables, such as education level, gender, and the country of residence, on this relationship. Finally, the author conducted hypothesis verification, confirming that this relationship can be expressed by a quadratic function, and then derived a quadratic function formula containing four fixed constants to represent the relationship between age and salary. This study concludes that the relationship between age and salary is best expressed as a quadratic function, influenced by factors such as education, gender, and country of residence.

Keywords: Linear regression; Non-linear regression; Age; Income.

1. Introduction

Currently, the term "economic downturn" has become increasingly popular in Chinese social media. Many working adults are saying that it is getting harder and harder to earn money. Therefore, the purpose of this study is to explore the impact of age and other social factors on wages [1]. Regarding the research on the relationship between age and salary, there has long been a debate in the academic community between linear and non-linear theories. The founder of human capital theory, Becker pointed out in his early research that as work experience and skills accumulated, the salary level would show a stable linear growth trend. However, this view was challenged in subsequent studies. Mincer conducted an empirical analysis and found that salary growth was not a continuous increase but reached a peak in the middle age and then gradually declined, forming an inverted U-shaped curve. This non-linear pattern was further verified in the cross-national data of the OECD, especially when analyzing the labor markets of high-income countries, it was particularly evident.

The influence of different countries' social and economic backgrounds on the age-wage curve is particularly prominent. Taking European countries as an example, researchers found that well-developed welfare systems and policies for delaying retirement have led to a generally later peak age for wages in countries like Sweden and Germany [2]. In contrast, economies in East Asia such as China and South Korea face the risk of wage stagnation or even decline for workers before the age of 40 due to intense competition in the labor market and the transformation of the occupational structure. Educational level is another key variable [3]. People also found through comparing the income data of different educational groups that higher education not only significantly increases the wage level but also mitigates the negative impact of age on income - the wage curve of highly educated individuals is gentler, and the unemployment risk is lower. Moreover, gender differences add complexity to this study. Some people pointed out that due to the interruption of career development or the occurrence of implicit discrimination during childbirth, the peak of the wage curve for women is often lower and the downward trend is more obvious [4].

Although existing studies have conducted analyses from perspectives such as the country, education, and gender, most of the literature only focuses on a single factor and lacks systematic exploration of the interaction effects of multiple variables. The innovation of this paper lies in

integrating these three types of factors and constructing a unified mathematical model to more comprehensively explain the relationship between age and salary. In this paper, the author mainly focuses on the relationship between age and wages, as well as the influence of factors such as education, the country where one lives, and gender. The final result shows that education and the country have a significant impact on a person's wages. For example, wages in some countries in the United States and Europe are generally higher than those in the three East Asian countries of China, Japan, and South Korea. Moreover, the rights of workers are better protected in the United States and European countries, while workers in East Asian countries are currently experiencing difficulties such as being laid off in middle age and having difficulty finding jobs after middle age. Section 2 of the paper will introduce some of the findings from previous studies in this area and will also introduce the mathematical models and methods that will be used in this paper. Section 3 will begin to explore other factors that affect wages and calculate the function that can express the relationship between age and wages.

2. Data and Method

For the data on the impact of educational attainment on wages in various countries, the author calculates an average or takes the median to summarize the statistics. For instance, when the wealth gap in a country is overly large, the author uses the median as the country's data. Conversely, for developed countries with relatively smaller wealth gaps, the author can use the average of the data as the reference figure for that country [5].

People's intuition generally holds that there are two possible relationships between age and salary. The first is that as one gets older, one's experience and ability also improve, and thus salary increases with age, presenting a linear relationship. Secondly, while salary increases with age during early career stages, it often plateaus and declines after middle age due to factors such as workforce reductions (layoffs) and diminished physical capacity to cope with work demands, resulting in an inverted U-shaped relationship. Therefore, one can make an intuitive assumption that the relationship between age and salary can be represented by a linear function representing a linear relationship or a quadratic function representing a non-linear relationship and use real data to calculate the P-value to verify which function relationship is feasible.

Since the author ultimately aims to derive a formula that can represent the age and salary in China mainland, all that is required is to calculate this formula first. Then, using some real-life examples, one can substitute the age and educational background into the formula to obtain a value. By comparing this value with the original one, the error of the formula can be observed.

3. Exploring Influence of Age and Various Social Factors on Income

3.1. Cross-Country Comparison

The author calculated the median and average values of the information found on internet websites to clearly contrast the differences in specific data regarding income peaks and peak ages across different countries in Europe, North America and East Asia.

Table 1. Comparison of peak age and peak age for different countries.

Country	Peak age (years)	Peak wage (USD)	Wage at the age (USD)
USA	47.5	85000	72250
Germany	50	78000	68640
Sweden	52.5	62000	57040
Japan	42.5	58000	39440
South Korea	40	50000	30000
China	37.5	35000	19250

First, the average income varies from country to country. From the statistical data shown in Table 1, one can observe several characteristics. For instance, the average income in the United States has reached the highest level, as its core industries are technology and finance. On the other extreme, the peak income age in China is the lowest. This is because many Chinese people are currently experiencing situations such as being laid off during their middle age. On the contrary, the peak income age in Sweden has reached an astonishing 52.5 years. Sweden's generous welfare system and pension incentives encourage people to delay retirement, while strong labor market protections and low age discrimination ensure that older workers can continue to accumulate earnings, pushing the peak income age even higher [6].

3.2. Education Impact

By organizing the information and calculating accurate data, the author created Table 2. It uses a high school diploma as 100% to calculate the salaries and unemployment rates for other educational qualifications. People with a master's degree or higher earn salaries that can even be three times that of those with only a high school diploma even after the age of 50. Moreover, the higher the education level, the lower the probability of career advancement at the age of 50.

Table 2. Comparison of the salary between different educational backgrounds.

Education Background/age	High school	Bachelors	Master & Doctoral
25-50(income)	100%	165%	203%
35-50(peak salary)	100%	189%	278%
After 50(income)	100%	211%	297%
After 50(unemployment rate)	100%	50%	32%

3.3. Impact of Gender on Wages

The author gathered some data to explore the influence of gender on work and salary. However, unlike other social factors, it is challenging to calculate the specific numerical value of the wage gap caused by gender.

Firstly, the World Bank 2020 report indicates that globally, women typically take a break from their careers for about 4-8 years due to child-rearing. But in countries with strong traditional gender roles (such as South Korea and Italy), this period can reach 10 years. Secondly, The OECD's "Gender Wage Gap Report (2021)" shows that the decline in female labor participation during the child-rearing period directly leads to their lifetime income being 20-30% lower than that of men. Thirdly, data from the Italian National Institute of Statistics (ISTAT) indicates that the income of women over 50 has decreased by 37%, mainly due to career interruptions during the child-rearing period and a high proportion of part-time work [7]. Fourthly, the research by the Korean Statistical Office (KOSTAT) shows that the employment rate of women aged 30-40 is 25 percentage points lower than that of men, directly affecting their later promotions and pensions. Fifthly, the ILO report (2022) states that in countries where the "men support the family" culture is deeply rooted (such as Japan, South Korea, and Southern European countries), the probability of women exiting the labor market after marriage is three times that of Nordic countries. Last point, the cross-national research by the Department of Economics at Harvard University found that Italian women's income decreased by 40% after childbirth, partly due to implicit discrimination by employers regarding "motherhood status" [8].

To conclude, women suffer from "workplace discrimination" due to traditional cultural beliefs in certain regions, and their wages are generally lower than those of men due to the demands of raising children.

3.4. Formula for the Relationship Between Age and Salary

To begin with using 10 real data points to create two subplots in Fig. 1, one for the linear function and the other for the quadratic function, as hypothesized by this work.

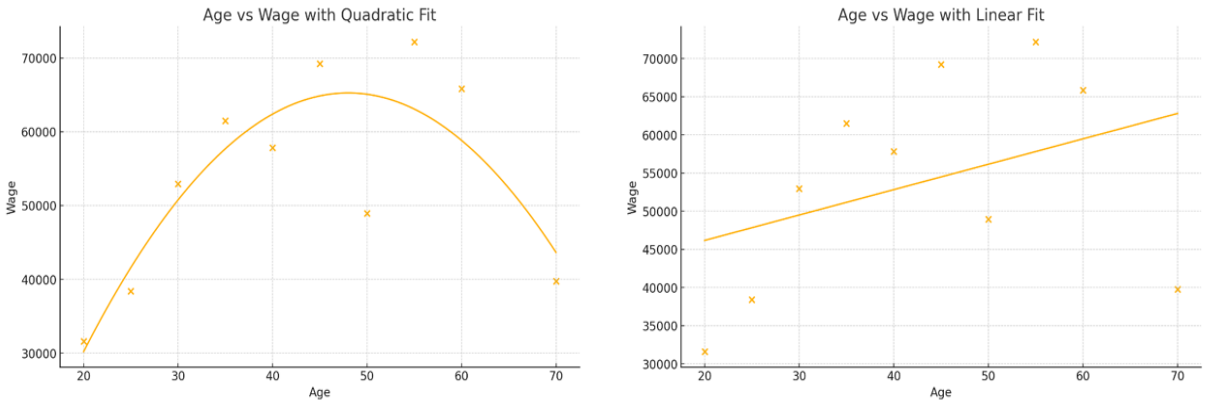

Fig 1. Express the comparison of this data using linear functions and quadratic functions. (Picture credit: Original).

Suppose this relationship is not a linear function. After calculation, all p-values are found to be greater than 0.05, which supports the hypothesis that the relationship cannot be expressed by a linear function. Conversely, if people suppose the relationship is not a quadratic function, the calculated p-values are almost all less than 0.05, leading to the rejection of this hypothesis and indicating that the relationship can indeed be expressed by a quadratic function. Then, the author constructed a quadratic function and substituted the real examples [9]. He had found into this function to obtain a set of formulas with the minimum error. In the basic formula, there are three constant terms that are known, namely the peak average wage of the country, the age at which the average wage reaches its peak, and a constant β_2 (β_2 is determined based on the rate of wage decline after reaching its peak in different countries.). For those with a high school education or above, their wages only need to be multiplied by the corresponding coefficient. In China, the peak wage is approximately 251,000 CNY (35,000 USD), occurring at around 37.5 years of age, with a constant value $\beta_2 = -514.5$. Note that this coefficient is applicable only when the monetary unit used in the formula is CNY.

Table 3. Comparison of actual salary data and predicted salary data.

Age	Education	Actual Salary (CNY)	Predicted Salary (CNY)	Error (%)
25	High school	52,500	525,00	0%
25	Bachelor's	120,000	99,225	-17.3%
30	Bachelor's	200,000	187,000	-6.5%
35	Master's	375,000	368,000	-1.9%
40	Bachelor's	275,000	260,000	-5.5%
50	High school	80,000	75,000	-6.3%
50	Bachelor's	400,000	377,000	-5.8%
65	High school	100,000	105,000	5.0%
65	Bachelor's	125,000	125,000	0%
65	Master's	225,000	225,000	0%

The comparison of actual salary data and predicted salary data is shown in Table 3. The reason why the coefficient multiplied by the educational level in this formula is based on the high school education level is 35 years old to 50 years old, and the coefficient of the highest wage difference during this period is chosen, is that when calculating the impact of education on wages, the author through calculation found that using the highest wage during the age range of 35 to 50 as the coefficient of the final formula can achieve the lowest error [10]. The fitting formula is as follows:

$$W(A,E) = \begin{cases} \frac{W_{peak}}{1.89} + \beta_2 (A - A_{peak})^2 & \text{High school} \\ 1.89\left[\frac{W_{peak}}{1.89} + \beta_2 (A - A_{peak})^2\right] & \text{Bachelors} \\ 2.78\left[\frac{W_{peak}}{1.89} + \beta_2 (A - A_{peak})^2\right] & \text{Masters/Doctoral} \end{cases} \quad (1)$$

To verify whether this formula is correct, the author prepared Fig. 2 to show the discrepancies and errors between the actual data and the data calculated by this formula. It is found that the two curves are rather close in value.

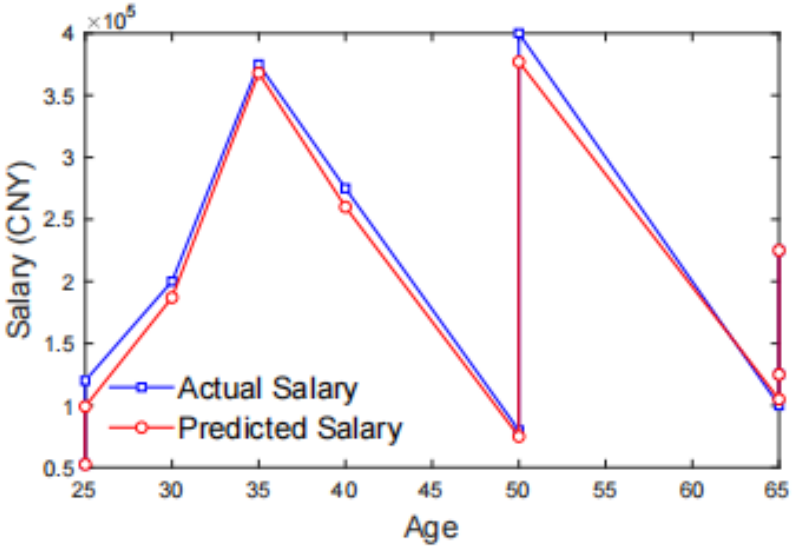

Fig 2. Comparison of actual and predicted salary.
(Picture credit: Original)

4. Conclusion

This study found that the relationship between age and salary can be expressed by a quadratic function. However, in the formula, additional constants such as educational attainment, the peak wage and peak age of the country where one is located (with different b values for each country), and a constant b (which varies for each country) need to be included to make the derived formula more accurate and cover more factors. It is clearly observable that educational attainment has an impact on salary. Even at the same age, the salary of a person with a doctoral degree can be three times that of a person with a high school degree. This research proves the importance and assistance that educational attainment brings to one's future. Additionally, this study also reveals the differences in work benefits and income gaps among different countries. The United States is the country with the highest salary among the six countries studied, and the factors include its main industries being technology and finance. Workers in European countries have the best treatment because their peak age is the highest. This is because compared to East Asia, the population density in Europe is lower, labor costs are higher, and workers in Europe need better rights. Compared to workers in other countries, Chinese workers have lower salaries and may face the problem of being laid off in middle age. This is because China has a large population density, so labor costs are cheaper, and China is still between developing and developed countries. This research objectively assesses the factors that affect salary, which is beneficial for people to understand the advantages and disadvantages of working in different countries and to recognize the importance of education for people. Finally, the author intends to continue researching what other factors affect salary in the future and will attempt to calculate a constant that can express the influence of gender on salary.

Reference

[1] Chen Xingxing. Research on influencing factors of domestic tourism consumption based on multiple linear regression analysis. Advances in Social Science, 2020, 9(9): 1530-1536.

[2] Li Wenyue, Li Jungang. Based on linear regression model, this paper makes an empirical analysis of China's GDP growth. Statistics and Application, 2022, 12(29): 1499.

[3] Song LZ. Financial support for rural development research: based on a nonlinear regression model. Shang. 2016;(8):141-142.

[4] Li Jun. Prediction method using the multivariate linear regression model based on the factor analysis and its applications on the share price forecast. CAJEPH, 2014, 5(30): 1-40.

[5] Sano Y, Mammen S, Houghten M. Well-Being and Stability among Low-income Families: A 10-Year Review of Research. J Fam Econ Issues. 2021, 42(Suppl 1):107-117.

[6] Walker CC, Druckman A, Jackson T. Welfare systems without economic growth: a review of the challenges and next steps for the field. Ecol Econ. 2021, 186:107066.

[7] Zhang L. and Wang Y. Wage differentials by education level in China's urban labor market. Chinese Economic Review, 2023, 75(3): 101-115.

[8] Card D. The causal effect of education on earnings. In: Handbook of Labor Economics. Vol 3. Elsevier; 1999:1801-1863.

[9] Blau FD, Kahn LM. The gender wage gap: extent, trends, and explanations. J Econ Lit. 2017, 55(3):789-865.

[10] Acemoglu D, Restrepo P. Tasks, automation, and the rise in US wage inequality. Econometrica. 2022, 90(5):1973-2016.