The 6th International Conference on Economic Management and Green Development (ICEMGD 2022)
DOI: 10.54254/2754-1169/3/2022813

Age and Hourly Wage: How Aging affect Earning

Mingze Li[1,a,*]

[1]Boston University, Boston, MA 02215, United States
a. mingzeli@bu.edu
**corresponding author*

Abstract: In an aging society, people need to focus on the consequence of aging in the labor market. This paper aims to uncover the effect of aging on employees' wages. To excluded the impact of the COVID-19 and inflation started at the beginning of 2020, the study is based on Current Population Survey (CPS) data collected in 2019. The author explored the issue by adopting the change in people's age and their hourly wage in the US labor market. The study introduced several regression models, and the analysis of the data shows a positive relationship between people's age and hourly wage. The findings also suggest that, overall, there is a tendency to raise hourly wage when people's age increase. In addition, the condition of being White, Asian, or married has a positive effect on people's hourly wage, while being Black, female, or disabled tend to have adverse effects on their hourly wage.

Keywords: labor market, regression models, hourly wage, racism

1. Introduction

Like many other parts of the world, with a declining birth rate, one of the most significant challenges in the U.S. labor market is aging. In a previous paper, Demand for Older Workers by Steven G. Allen from North Carolina State University demonstrates that the share of the U.S. labor force consisting of workers with age of 55 and older has risen from 12.4 percent to 23.1 percent from 1998 to 2018 [1]. As the number of older age workers becomes even more significant, society cannot ignore the aging issues.

The paper compares age and people's hourly wage instead of age and salary since hourly wage is a precise measurement when take into consideration of some people's ability to work decreases as age increases [2], which can have an adverse effect on people's wages since the speed of wage increase is not on the same pace as the speed of aging [3]. In the paper, the author uses annual data extracted from IPUMS - CPS of 2019 [4] to create econometric models on the relationship between age and hourly wage. The data was specifically chosen to avoid the impact of the COVID-19 and inflation on each individual's wage and its relationship with age. In addition, the author adds six variables to the study to reduce the endogeneity concerns in all models and incorporate possible explanations for changes in hourly wage other than effect of age.

In the previous study done by Myck and Michał in 2007 [5], they conducted a correlation of age and wage across Germany and British by using collected samples of male workers from the year 1995 to 2004. Their study shows respondents' wages tend to increase while they are aging. This paper aims to test if their conclusion on hourly wage and age is still applicable to the U.S. labor market.

DOI: 10.54254/2754-1169/3/2022813

In order to make the best fit of the data, the paper conducted both simple and multiple regression models to analyze the relationship. To be more detailed, this research introduces six regression models. For signal variable regression, the study includes Simple Regression Model, Quadratic Regression Model, and Log-Linear Regression Model. For the Multiple Regression, the paper includes the Multivariable regression model, Log-Linear Multivariable regression model, and Multivariable regression model with an interaction term. The author realizes it is essential to contain models that conduct different forms of transformations to fit the data, a Log-linear model, and a model with an interaction term to study the marginal effect of one independent variable over another.

2. Descriptive data and data cleaning methods

2.1. Survey Data.

The data in the paper is collected from the Current Population Survey, CPS, of annual data collected in 2019. The U.S. Census Bureau administrates the CPS data for the Bureau of Labor Statistics in the U.S. labor market. The data includes a probability selected sample of 60,000 households across all U.S. Census Bureau field representatives document 50 states and the District of Columbia, all samples through both personal and telephone interviews. In the survey, all respondents are 15-year-old and above, while respondents in institutions are not included, such as prisons and long-term care hospitals. The study examine the relationship between age and hourly wage, to exclude the effect of COVID-19 and the large-scale worldwide inflation while using the most up-to-date information, the author decided to conduct the study by adopting 2019 data, which contains 180,101 responses.

2.2. Definition of Variables.

In the study, in addition to age and hourly wage, six other variables that might have an effect on people's hourly wage were analyzed since many scholars have previously pointed out more factors might have additional effect the dependent variable [6]. All data are from the Current Population Survey in one data set .

2.2.1. Age.

It was collected on each person's age on the last birthday. In the study, age between 80-84 is collected as 80, and age 85+ is collected as 85.

2.2.2. Gender/Sex.

It is a binary variable that indicates each person's gender. In the 2019 data set, there are 87,336 males and 92,765 females.

2.2.3. Race.

The race variable includes 26 categories of all respondents' races, such as White, Black, American Indian, White-Black, Asian or Pacific Islander, White-Asian, etc. In the study, three races racial variables were selected that make up most of the races, White (139,006), Black (21,141), and Asian only (11,226).

2.2.4. Marital Status.

This variable describes each respondent's current marital status, including whether the spouse was currently living in a same household. The variable, Marital Status, includes six categories of matrimonial conditions, Spouse present, Married, Married, Spouse absent, Separate, Divorced,

DOI: 10.54254/2754-1169/3/2022813

Widowed, and Never married/single. In the study, the author only focused on whether the person is married, so I treated all other conditions other than married as un-married.

2.2.5. Education.

This variable introduces the education level of all respondents and is measured by the highest year of education/degree completed. The category of education includes a wide range of samples from Grades 1, 2, 3, or 4 to Doctorate degrees.

2.2.6. Work Disability.

The work disability variable indicates whether the respondent has a disability that prevents the respondents from working or limits the types or amount of work.

2.2.7. Hourly Wage.

This is the dependent variable in the study, and it is collected by documenting how much a respondent earns in each working hour of the current job. In the data set, there are some respondents who are documented with an hourly wage of 999.99, which means missing data.

2.3. Methodology for Cleaning Data

The survey data from the Current Population Survey follows the standard format, which contains missing data, and the data must be cleaned while converting them to the correct format for the study. For example, any missing value needs to be removed or a new variable created based on an existing variable to fit the study. Therefore, needs to be cleaned into the format that is needed to fit the econometrics model in the study.

2.3.1. Gender/Sex.

It is a categorical variable, and the study wants to focus on how being a female effect the hourly wage. In the data cleaning, the author assigns all females as 1 and males as 0.

2.3.2. Race.

The race variable includes 26 different categories, and the study focus on the effect of the three most prominent races in the sample, which are White, Black, and Asian. To distinguish the three races' categorical variables, instead of creating the signal variable, the author constructed three categorical variables, White, Black, and Asian, to represent each race group.

2.3.3. Marital Status.

This variable includes four categories, and the article will investigate if the married status will affect people’s hourly wage, so the categorical variable is constructed, married, by assigning 1 as married and 0 for another status.

2.3.4. Education.

In the CPS data, education has always been labeled in unique codes that match different years of education. In order to covert the raw data to fit the model, the class year was used to represent the level of education by replacing the original code. For instance, if the class year is 12, it means the respondent receives a high school degree.

DOI: 10.54254/2754-1169/3/2022813

Table 1: The Table of effect of Education Level on Hourly Wage.

VARIABLES	(1) Non-degree	(2) High School Degree	(3) Associate Degree	(4) College Degree	(5) Master's Degree	(6) Doctorate Degree
Age	0.116***	0.119***	0.115***	0.115***	0.105***	0.112***
	(0.00822)	(0.00818)	(0.00822)	(0.00809)	(0.00813)	(0.00811)
Non-degree	-6.093**					
	(2.444)					
White	0.882	0.843	0.880	0.575	0.717	0.815
	(0.644)	(0.639)	(0.644)	(0.633)	(0.634)	(0.635)
Black	-1.093	-1.021	-1.095	-1.236*	-1.183*	-1.131
	(0.720)	(0.716)	(0.720)	(0.708)	(0.709)	(0.710)
Asian	3.986***	3.644***	3.990***	2.803***	3.517***	3.555***
	(0.832)	(0.827)	(0.832)	(0.821)	(0.820)	(0.821)
Female	-2.934***	-3.144***	-2.963***	-3.196***	-3.042***	-3.032***
	(0.235)	(0.235)	(0.235)	(0.232)	(0.232)	(0.232)
Married	2.939***	2.877***	2.896***	2.716***	2.784***	2.893***
	(0.253)	(0.251)	(0.253)	(0.249)	(0.249)	(0.249)
Disable	-1.282**	-1.257*	-1.296**	-1.248**	-1.205*	-1.153*
	(0.647)	(0.642)	(0.646)	(0.636)	(0.636)	(0.637)
Highschool		-2.519***				
		(0.248)				
Associate			1.090***			
			(0.359)			
College				4.935***		
				(0.308)		
Master					8.752***	
					(0.565)	
Doctorate						18.25***
						(1.28)
Constant	12.98***	13.85***	12.90***	12.71***	13.31***	13.13***
	(0.698)	(0.699)	(0.699)	(0.687)	(0.688)	(0.68)
Observation	7,359	7,359	7,359	7,359	7,359	7,359
R-squared	0.093	0.105	0.093	0.123	0.121	0.118

2.3.5. Work Disability.

To differentiate people who are disabled, the author assigns 1 to disabled people and 0 to all no disabled respondents.

DOI: 10.54254/2754-1169/3/2022813

2.3.6. Hourly Wage

The dependent variable hourly wage is a continuous variable with missing data when the value of the variable is 999.99. In this case, Ii is only necessary to remove the effect of the missing data by dropping all the 999.99/missing data. There are 172,742 data dropped from the data set, which results from 7,359 valid responses in the data set.

2.4. Data Attrition

Nowadays, in addition to aging, education level acts as a significant role in affecting people's earnings [7]. This paper also conducts the effect on hourly wage if people possess different education degrees, including no degree, high school degree, associate degree, college degree (4 years), master's degree, and doctorate degree. Therefore, based on the cleaned data set, the respondent who possesses no educational degree is about 0.23% of the total respondent, high school degree holders take approximately 33.67%, respondents who possess an associate degree is about 12.27%. In comparison, a college degree and above takes up approximately 22.56%, with 4.44% being master's degree holders and 0.88% being doctorate or professional degree holders of the total sample size.

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

In the table, the effect of education level on an hourly wage, the level of education acts as a significant factor that has a decisive impact on a people's income. By possessing a high school degree and below, the coefficient of variable, education level, is negative while people with at least an associate degree (and above) tend to have a positive coefficient of education level, which has proven to be statistically significant when possessing a different educational degree. The pattern of the table also shows companies nowadays are willing to pay much more for an employee with a master or doctoral degree.

According to data of the six levels of education, when respondents have no educational degree, the coefficient of education level is -6.093, which, on average, makes $24.34 per hour less compared to people who have a doctoral degree. At the same time, on average, college graduate makes $3.845 per hour more than people who have only an associate degree.

3. Econometric models

At a fundamental level, when exclude the effect of multivariable on both response and explanatory variables, the objective of the model is to examine whether older workers are receiving higher salaries while aging. The univariable simple regression model serves as a foundation of multivariable modeling with more variables and various functional forms. The study will also analyze simple regression models, which will include analysis over multiple regression model by adding additional independent variables and variable transformation, including education level, race, marital status, gender, and disability.

3.1. Simple Regression Models:

$$Hourwage = \beta_0 + \beta_1 * age + e^1 \quad (1)$$

$$Hourwage = \beta_0 + \beta_1 * \text{age}^2 + e^1 \quad (2)$$

$$LogHourwage = \beta_0 + \beta_1 * age + e^1 \quad (3)$$

DOI: 10.54254/2754-1169/3/2022813

The study of models will begin by analyzing a simple linear regression models. Workers' age will be adopted as an independent variable, X, while hourly wage as a dependent, Y, variable to inspect how aging influence their income measured by hourly wage. The coefficient β_1 represents the change in average hourly wage, on average, when age ascent by one year, while coefficient β_0 shows the hourly wage, on average, received by a worker at the age of 15 since the age is begin measured at the age of 15, which is when the value of age variable equals 0. The two values of the coefficient are parameters of the population thus unknown, so sample data analysis is needed to estimate the parameters.

The second function is the quadratic regression model. Like the simple linear models, the quadratic regression models also include only two quantitative variables, age, and hourly wage. However, different from the simple linear model, β_1 the quadratic regression model shows the change in hourly wage, on average, when the variable, square of age, raise by 1 unit, while the derivative can be found. The derivative of the model with respect to age is equal to 2 * age * β_1, which shows that when age increase by one year, the average hourly wage increases by 2 times age * β_1. The predictions of the quadratic regression model shows that it will provide the lowest SSE, which is the most effective model in capturing the relationship between hourly wage and age.

The histogram of hourly wage is skewed to the right, which shows the simple linear regression model does not follow the normality assumption in four OLS assumptions. Therefore, I manipulate the data by adding them into a more symmetric distribution through a logarithmic transformation, in which I construct a Log-Linear regression model to fit the data. Log-Linear regression model contain a coefficient, β_1, which describes the change in log of an hourly wage, on average, when age raise by one year. The slope of the model represent by β_1 * hourly wage, which reveals that when the of age increases by one unit, the log form of hourly wage increases by β_1 * age US dollar.

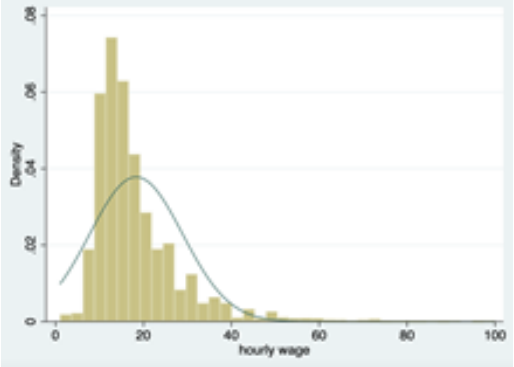

Figure 1: histogram of hourly wage.

3.2. Multiple Regression Models:

$$Hourwage = \beta_0 + \beta_1 * age + \beta_2 * educ_{rev} + \beta_3 * white + \beta_4 * black + \beta_5 * asian + \beta_6 * female + \beta_7 * married + \beta_8 * disable + e^1 \quad (4)$$

$$\log hourwage = \beta_0 + \beta_1 * age + \beta_2 * educ_{rev} + \beta_3 * white + \beta_4 * black + \beta_5 * asian + \beta_6 * female + \beta_7 * married + \beta_8 * disable + e^1 \quad (5)$$

DOI: 10.54254/2754-1169/3/2022813

The purpose of introducing multivariable regression models is to incorporate additional possible explanatory variables that could have an additional effect on the hourly wage of the respondent other than the effect of aging. Similar to the simple regression model, the author includes a Multiple Linear and Log-Linear multiple in the multiple regression model section. The goal of having the logarithmic transformation is to construct a better fit for the data under the four OLS assumptions. The multivariable regression models include two quantitative variables, educ_rev and age, and six categorical variables (white, Asian, black, female, married, disabled). The construction of multivariable regression models could promote the study reveal if variable, age, is the most significant factor in determining a respondent's hourly wage, which is also the lies in the previous expectation.

3.3. Multiple Regression with Interaction Term:

$$\begin{aligned} Hourwage &= \beta_0 + \beta_1 * \text{age} + \beta_2 * educ_{rev} + \beta_3 * white \\ &+ \beta_4 * black + \beta_5 * asian + \beta_6 * female + \\ &\beta_7 * married + \beta_8 * disable + \beta_9 * age\_ + e^1 \end{aligned} \tag{6}$$

This model of multiple regression with interaction terms is created to examine whether the marginal age effect on hourly wage depends on people's education level, and the marginal effect of education level on hourly wage depends on the people's age. This paper hypothesize that respondents of higher age are more vulnerable to encounter various issue caused by their physical conditions, so they have reduced productivity and experience a diminishing hourly wage as aging. Therefore, creating the interaction term within the model can help to explain this effect.

4. Results

This article conducts six econometric models. Models 1 to 3 only study the direct effect of age on hourly wage, while the results show age and hourly wage are positively related. Model 4 adds more variables that also affect hourly wage. Coefficients of age in models 4, 5, and 6 also show a positive relationship between age and hourly wage. However, additional variables in models 4, 5, and 6 can have negative coefficients since some factors are negatively related to hourly wage; for instance, coefficients for Black are negative in all three models because Black employees often encounter unfairness when entering the U.S. jobs market.

4.1. The Simple Linear Regression model

Model 1 shows a positive relationship between age and hourly wage. The simple linear regression only one independent variable, age, and one dependent variable, hourly wage. The coefficient of age is 0.142, which shows when age increase by one year, people's hourly wage increase by $0.142. Then OLS assumption was tested to examine whether the simple linear line can be used as the best model to fit by predicting the variable and constructing the scatter plot graph with age(Figure 2). The scatterplot of simple linear regression line shows the heteroskedasticity relationship between the age and hourly wage, which indicates the simple linear line is not the best model to describe the relationship.

However, the simple linear regression model only has an R-square of 0.043, which explain 4.3% of the variation in the model, and other factors explain 95.7% of the variation in hourly wage, which is considered a very low variation. In order to explain other variations in the model, the multiple regression models were introduced in the study.

DOI: 10.54254/2754-1169/3/2022813

Table 2: Regression table for six models.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Model1	Model2	Model3	Model4	Model5	Model6
age	0.142***		0.00758***	0.103***	0.00570***	0.188***
	(0.00786)		(0.000354)	(0.00775)	(0.000348)	(0.0400)
educ_rev				1.462***	0.0640***	1.749***
				(0.0471)	(0.00212)	(0.140)
white				0.567	0.0385	0.559
				(0.606)	(0.0272)	(0.605)
black				-1.086	-0.0259	-1.107
				(0.677)	(0.0304)	(0.677)
asian				2.254***	0.0890**	2.209***
				(0.785)	(0.0352)	(0.785)
female				-3.676***	-0.185***	-3.693***
				(0.223)	(0.01000)	(0.223)
married				2.394***	0.125***	2.368***
				(0.239)	(0.0107)	(0.239)
disable				-1.257**	-0.0954***	-1.265**
				(0.608)	(0.0273)	(0.608)
age2		0.00135***				
		(9.07e-05)				
age_educ						-0.00651**
						(0.00300)
Constant	12.53***	15.77***	2.477***	-4.829***	1.714***	-8.537***
	(0.342)	(0.211)	(0.0154)	(0.872)	(0.0392)	(1.919)
Observations	7,359	7,359	7,359	7,359	7,359	7,359
R-squared	0.043	0.029	0.059	0.197	0.216	0.198

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

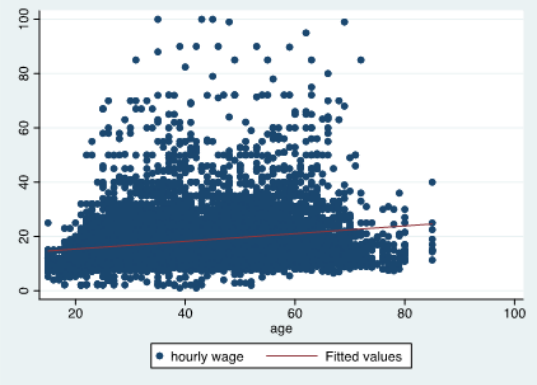

Figure 2: Simple Linear Regression Line.

DOI: 10.54254/2754-1169/3/2022813

4.2. The Multiple Log-Linear Regression Model

Model 5 has an R-square of 0.216, which account for the 21.6% variation in the model. Like the multiple linear model, the multiple log-linear model has the same independent variables, and the dependent variable is converted to a log form (hour wage). Compared to multiple linear models, the primary purpose to incorporate a multiple log-linear model is to investigate whether it fits the data better than the multiple linear model.

The Log Linear multiple regression model shows two factors that are influential to people's hourly wage, which are gender and race. The econometric model has shown that an individual who is Asian or White is more likely to earn a relatively higher income in his/her workplace by $0.038 per hour for White and $0.088 per hour for Asians. At the same time, a Black tends to get paid less by $0.025 per hour, potentially because of cultural differences or discrimination. To better investigate the racial effect on hourly wage, the three categorical variables of white, Asian, and black are added to the model. The White/Black/Asian variable will equal 1 if an observation is indicated as White/Black/Asian, and the variable will equal 0 if otherwise. Like race variable, female is also added to the model as a categorical variable to investigate the potential effect of gender on hourly wage, and the study returns that when holding all other variables constant, being female will lead to a $0.184 per hour decrease, on average, compared to male. The study shows income vary between males and females in the United States since male are often have easier access to education, while they are more likely to be promoted to influential positions in the work setting. The female variable equals 1 when respondent are female and 0 for male respondents.

Marital status is another qualitative variables that may play a important role in explaining people's hourly wage differences. The coefficient for married is 0.125, which means that married people tend to make $0.125 per hour more compared to not married people. One explanation is that married people tend to have higher work motivation which they are willing to hard on their current jobs to improve the standard of living. In contrast, single individuals, divorced, or widowed are less encouraged. The research by May Luong and Benoit-Paul Hebert (2015) indicates that handicapped employees with specific disabilities are less likely to access well-paid jobs because of a lack of job options due physical limitations [8]. The Log-Linear multiple regression returns those disabled workers are likely to make less by $0.095 per hour when holding other variables constant, compared to no disabled workers.

Taking in to consideration of all factors, the study conclude a thorough understanding of the correlation between hourly wage and people's age. The author can determine whether age is the most crucial variable that impact the response variable, or some of the other independent variables acts a more influential role. However, in the model, there could still be lurking variables, which affect both hourly wage and age in this model. Therefore, This study also examine the underlying correlation between the hourly wage and workers' age when including additional dependent variables.

The result proved that age affects people's hourly wage by $0.142 per hour when excluding all other variables. At the same time, when adding additional variables to the models, the relationship between age and hourly wage is still positive but with a lower slope of 0.005 since, in the model, other variables might have a more significant impact on an hourly wage. Additionally, all coefficients of age are positive across all six models, which proves that age and hourly wage are positively related. In other words, while people are aging, their hourly wages tend to increase simultaneously.

5. Conclusion

This study primarily focused the influence of respondents' age on their hourly wage received on their job. The research idea in this paper is inspired by previous studies done by Myck and Michał, which studied a similar topic of the relationship between wage and age using data collected in British and

DOI: 10.54254/2754-1169/3/2022813

Germany [5]. This article focuses on examining the correlation between age and hourly wage in the US labor market. While it is difficult to identify the "pure effect" of age on hourly wage, inspired by the study by Peña and Raúl De La includes additional variables, including race, education level, marital status, etc [9]. Overall, using hourly wage instead of annual income or weekly earnings is to avoid the effect of potential health issues and hidden effects that might affect elderly people in their working performance.

After cleaning the data, the author analyzes 7359 valid responses to fit into the models. The models selected in the study included signal variable linear, quadratic, log-linear, multiple-variables log-linear regression, and multiple-variables linear regression with interaction terms, which all of them reveal a related weak positive correlation between age and hourly wage. In other words, the study confirms people's hourly wage increases while they are aging. The study also reveals other conditions can present a minor impact on people's hourly wage, both positively and negatively. The finding shows conditions of being White, Asian, or married can have a positive effect on people's hourly wage while being Black, female, or disabled tend to have adverse effects on their hourly wage.

At the same time, the models in this study can never be perfect when they are used to predict the relationship between people's age and their hourly wage. It is believed that the results of the study can be refined by adding more sample data since the central limit theorem states a larger sample size will produce a normally distributed model while leading to a more accurate prediction of results, including standard deviation and mean. In the future investigation, scholars can include more independent variables, including the number of households or hours of working, to conduct a study on how other conditions affect people's hourly wage. Another valuable direction to study is to conduct how the correlation between aging and rate change of hourly wage led to improvement in public policy to support the group worker experience affected by aging.

References

[1] *Allen, S. (2019). Demand for older workers: What do economists think? what are firms doing? https://doi.org/10.3386/w26597*

[2] *Balderrama, C., Flores, J., & Maldonado, A.: The Effects of Age to Meet the Standard Production Rate. Work,vol. 51, no. 4, 2015, pp. 827–37. Crossref, doi:10.3233/wor-141972. (2015).*

[3] *Van Ours, J.C., Stoeldraijer, L.: Age, Wage and Productivity. SSRN Electronic Journal, https://doi.org/10.2139/ssrn.1563992. (2010).*

[4] *Flood, S., King, F., Rodgers, R., Ruggles, S., Warren, J.R.: Integrated Public Use Microdata Series, Current Population Survey: Version 8.0 [dataset]. Minneapolis, MN: IPUMS,https://doi.org/10.18128/D030.V8.0, (2020).*

[5] *Myck, M.: Wages and Ageing: Is There Evidence for the 'Inverse-U' Profile? Oxford Bulletin of Economics and Statistics, vol. 72, no. 3, pp. 282–306. Crossref, doi:10.1111/j.1468-0084.2009.00582.x. (2010).*

[6] *Freeman, R.B.: The Effect of Demographic Factors on Age-Earnings Profiles. The Journal of Human Resources, vol. 14, no. 3, p. 289. Crossref, doi:10.2307/145573, (1979).*

[7] *ZWANE, T.T.: The Causal Effect Of Education On Earnings In Urban And Rural South Africa: A Further Update. Demography and Social Economy, no. 1, pp. 79–94. Crossref, doi:10.15407/dse2020.01.079.(2020).*

[8] *Luong, M., Hébert, B.P.: Age and Earnings. Statistics Canada, www150.statcan.gc.ca/n1/pub/75-001-x/2009101/article/10779-eng.htm,(23 Jan. 2009).*

[9] *De LaPeña, R.: Jacob Mincer. Schooling, Experience and Earnings. Nueva York, [N. Y.]: Columbia University Press, 1974. 152 p. Estudios Demográficos y Urbanos, vol. 9, no. 01, p. 128. Crossref, doi:10.24201/edu.v9i01.316, (1975).*