

The Influencing Factors of Income in the UK under the COVID-19 Epidemic

Yutong Shu^{1,*}

¹Faculty of science, University of Strathclyde, Glasgow, UK

*Corresponding author: yutong.shu.2021@uni.strath.ac.uk

Abstract. In December 2019, there was a sudden outbreak of a highly contagious disease which is named coronavirus (COVID-19) in China. In just two months, the coronavirus has spread to nearly 50 countries. In 2020, the epidemic caused a large number of employees losing their jobs. This paper taking income as the main research object, because when the employees were unable to make profits for the companies or enterprises, many enterprises were living beyond their means and were on the verge of closing down. As a result, Lay-offs increased and employees' incomes were affected. After that, it is still necessary to collect data such as employment and education and other independent variables. This paper adopts linear regression and nonlinear regression to explore the relationship between income and other influencing factors. Besides that, the research will also focus on whether these data are independent and build a model between income and other influencing factors. Finally, this paper will draw a conclusion that income has a strong relationship with employment. After building a model, it also proves that crime also make a difference to the income.

Keywords: COVID-19, Income, Employment, Crime, Education.

1. Introduction

The purpose of this study is to find out the correlation between people's income and various influencing factors in a single country: the UK when it is affected by a wave of epidemic, and briefly comment on existing theories and established models.

The first confirmed case of acute respiratory infection due to coronavirus in the United Kingdom was detected in the UK on 31 January 2020. At the beginning, the measures that the government taken were to isolate high-risk groups and require citizens don't get out except for emergencies. According to Razai, the aim of all these policies is to control the spread of the epidemic [1]. According to Johnson, it is roughly predicted that the coronavirus pandemic worldwide will cause a more serious economic shock than the global financial crisis in 2008 [2]. From Adams-Prassl's paper, using data from a UK survey collected on 25 March 2020, shows that approximately 57% of employees are paid less than they were paid before the pandemic. About 8% lost their jobs in the previous month or so, and 33% of those still working had lost their jobs in the next four months. Overall, the average employee's salary was 35% lower than before the pandemic [3].

In addition, expect job losses and layoffs due to the pandemic, there are still many other reasons that affect the average income of employees. The epidemic has had a significant impact on the economy which also leads to a remarkable impact on the real estate market. In order to reduce the spread of the epidemic, economic activity has declined significantly, starting with the service and hospitality industries which are often exposed to risks due to the inability to meet requirements such as rent and labor costs. Secondly, from Kholodilin's paper, for private tenants, because of the epidemic control measures, this kind of impact will also happen to them [4]. Landlords are likely to raise the price of the remaining houses, resulting in an increase in the rent paid by individual tenants and a decrease in their income.

The third influential factor is the impact by education. Under the intervention of policy, many high education institutions in the UK chose to adopt online teaching, temporarily replacing offline teaching mode. This policy has had a huge cultural impact on the classic teaching model in Britain. Online teaching mode strictly inspects students' autonomous learning mode and self-discipline. According to Lydia Arnold, autonomous learners need to make decisions about their own learning [5]. As Fazey's opinion, autonomous learners need to be confident in learning and accepting new knowledge

[6], and according to Dickinson, autonomous learners must have the ability to learn actively and independently [7]. For schools with a large proportion of international students, some students may choose to stay in their own country rather than go to another country which is not familiar to study during the online teaching period, which may affect not only the income of students' families, but also schools and surrounding institutions.

At the same time, Health may also have a greater impact than other various factors which affecting income. According to Angus Deaton's paper, economists believe that high income is negatively correlated with various risky behaviors, such as alcoholism, smoking and obesity [8]. Individuals who abandon these unhealthy habits are likely to earn more and live in better health. On the other hand, people who are used to smoking and drinking may earn a lower income, while these risky behaviors also have an impact on their health. Although these are more biased towards the impact of income on health, it is also good evidence that there may be a link between income and health. In this wave of epidemic, people's health has received a very big threat. According to Zhang J, because of the coronavirus, many people come to the hospital in preparation for treatment of the coronavirus. At the same time, the number of patients with other diseases has not decreased, which also leads to hospital doctors become busier, more and more patients [9]. At the same time, the patient's health is generally worse than that of a healthy adult, so the possibility of hospital-acquired coronavirus infections in patients with other diseases is also higher. As a disease with a very high transmission rate and a very high mortality rate, many people in the aftermath of coronavirus are often absent from work or bedridden, and it is not just bad habits such as smoking and drinking that affect an individual's income.

Finally, there are influencing factors about crime. According to Patterson, only considering the crime types which are not very serious, poor areas are more correlated with crime rate [10]. At the same time, the areas where high-income people lived is less correlated with crime rate. However, according to Kim, after analyzing the data of more than 30 European countries, he came to the conclusion that the income inequality of different groups did not increase the crime rate in this region, because the crimes caused by income inequality only accounted for 3% of the crime results [11].

This paper aims to complete the research which is analyzing the above factors and finding whether there is a relationship between these relevant factors. Other influencing factors, such as living and housing, will also be analyzed in the study to determine whether these factors also have a significant impact on income.

2. Methodology

This study will analyze the research problems below. First of all, this paper will find the group of data with the largest correlation coefficient and strongest correlation with income among various influencing factors. This group of data will be recorded as the main independent variable, and linear regression between the dependent variable and the main independent variable will be conducted to obtain the relationship obtained after regression by the least square method. Secondly, this paper will use the log model and sqrt model for analysis, draw Q-Q plots and residual plot of these two models which are after transformation, and further determine the correlation between the dependent variable and the main independent variable. Then this paper will calculate the confidence interval and prediction interval for the given data, and add a 95% regression curve. The formulas of the confidence interval and the prediction interval is:

$$CI = \hat{y} \pm t_{(n-p-1, \alpha/2)} \sqrt{\sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SS_{xx}} \right)} \quad (1)$$

$$PI = \hat{y} \pm t_{(n-p-1, \alpha/2)} \sqrt{\sigma^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SS_{xx}} \right)} \quad (2)$$

In the two formulas above, $\hat{y}$ is the average of the dependent variable, σ^2 is the variance of the independent variable, n is the number of research objects participating in the experiment, x_0 is

each different independent variable, $\bar{x}$ is the average of independent variable. After calculating the confidence interval and the prediction interval, this paper will use backward Selection to select the most appropriate model. The principle of this method is: in the first modeling, all independent variables are added to the equation and the largest independent variable of p-value is removed from the obtained model, so as to deduce the class. Finally, when all p-values fall within the acceptable range, the final model is obtained. This model includes all the influential factors with the stronger correlation with dependent variables as independent variables. Assume that different independent variables are $X_1, X_2, \dots, X_n$. In addition, let the constant term of the regression equation be β_0 and the variance be ε . Last but not least let the coefficients before different independent variables in this regression equation be $\beta_1, \beta_2, \dots, \beta_n$. After optimizing the model with backward selection method, it can get a regression equation containing β_1, β_2 and other different parameters:

$$\text{Income} = \beta_1 X_1 + \beta_2 X_n + \dots + \beta_n X_n + \beta_0 + \varepsilon \quad (3)$$

Finally, the regression model between the dependent variable and multiple re-selected independent variables is established, and the coefficients before different independent variables are calculated to obtain the final regression model.

3. Results and discussion

3.1 The correlation coefficient and the regression equation

Firstly, the correlation coefficient between dependent variable income and independent variable employment is calculated. Correlation coefficient is a statistical indicator to study the degree of correlation between different variables. The larger the correlation coefficient is, the stronger the correlation between two groups of data is. R1 was used to represent the correlation coefficient, and the correlation coefficient R1 was 0.9608607. After calculating the correlation coefficient before other independent variables and income, it can be known that the correlation coefficient between income and employment is the largest. Therefore, employment can be selected as the independent variable, and then the income and employment can be regression, and the binary linear regression equation between them can be worked out.

The results obtained by regression can construct a binary linear regression equation $y = 0.9632x + 5.7624$ for income and employment. According to the regression equation, income is positively correlated with employment. As income increases, employment continues to increase. After the regression, in order to give enough evidence to prove that income is strongly associate with employment, and is there any advantageous evidence for the relationship between two groups of positive correlation data sets, this research can use log model and sqrt model and then draw the Q-Q plot and the residual plot.

3.2 Q-Q plots and residual plots

First, this research use log model to research. Assume that these two data sets are independent. According to figure 1, which is the Q-Q plot between employment and income, it can still be considered as a normal situation although the residuals look like normal distribution. In addition, use log (employment) as the abscissa and the residual as the ordinate to draw a residual plot. According to the figure 2, there is no evidence to prove that residuals and fit values is a randomly uniform distribution. As the depicted points are randomly distributed in the both side of the line with residual equal to 0, it indicates that the regression line fits the meta-observations well. Otherwise, it indicates that the fitting of the regression line to the original observation value is not ideal. From the figure 2 it is obvious that the variation trend of the residuals is first increase slightly and decrease after reach the maximum. This gives strong evidence to prove that the independence hypothesis probably not correct.

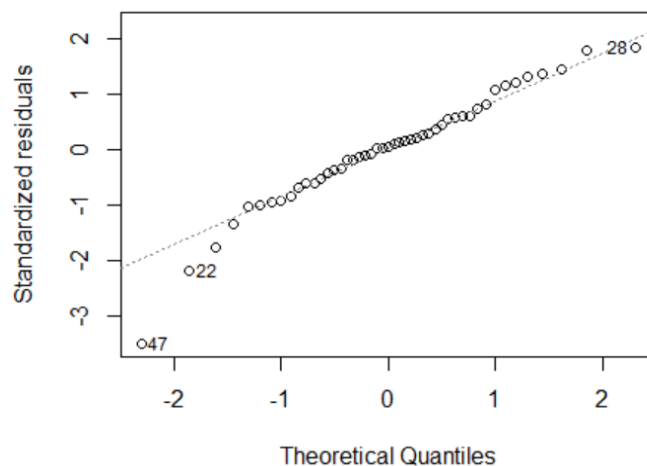


Figure 1. Q-Q plot of log (employment) and income

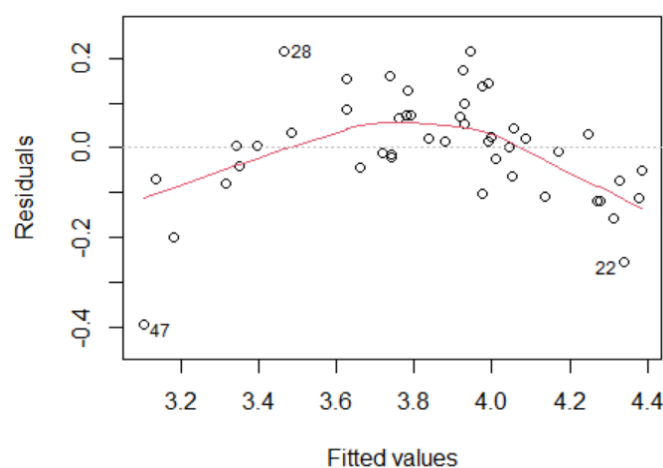


Figure 2. Residual plot of log (employment) and income

In addition, this research use sqrt model to research. According to the figure 3 which is the Q-Q plot between employment and income, the residuals are also look like normal distribution, so it can still be considered as a normal situation. At the same time, from the figure 4, the residuals' variation trend is increase steadily first and then decrease gradually. This also gives strong evidence to prove that the independence assumption probably not correct. As a result, there are enough evidences that there is a strong relationship between income and employment.

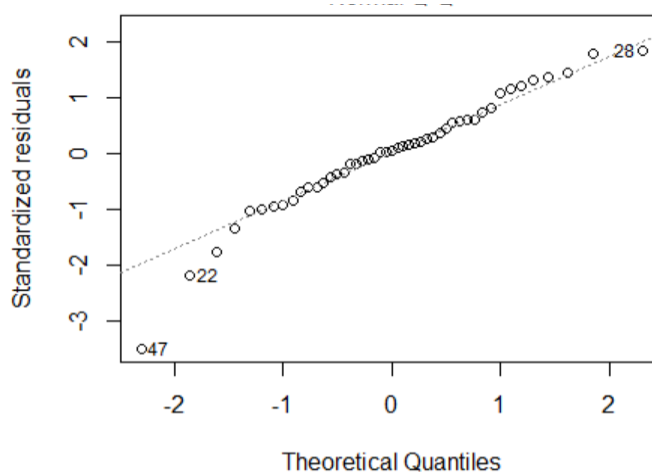


Figure 3. Q-Q plot of sqrt(employment) and income

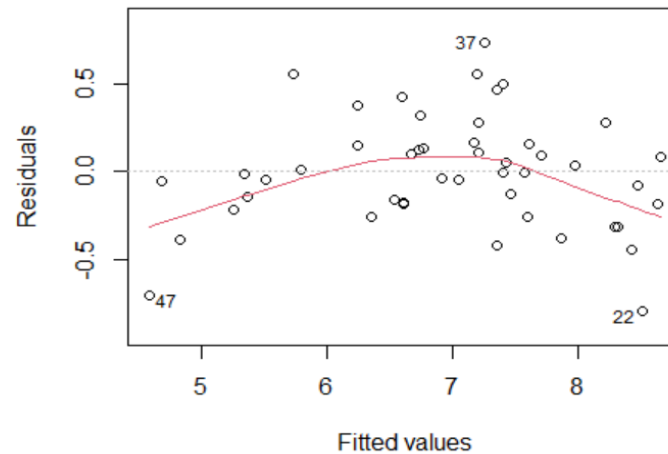


Figure 4. Residual plot of sqrt(employment) and income

3.3 The confidence interval and the prediction interval

Next, this paper will plot the prediction interval with the given data. After that, the 95% confidence interval will also be added in the plot. Firstly, using the average formula to calculate the average of income, which is 46.7636. Then calculating the RSS with this formula: $RSS = \sum (y_i - \hat{y})^2$. The result is that $RSS = 26131.12$. And then the σ^2 is $\frac{RSS}{n-p-1} = 218.544$. Then using formulas 1 and formulas 2 to find out the confidence interval and prediction interval as figure 5.

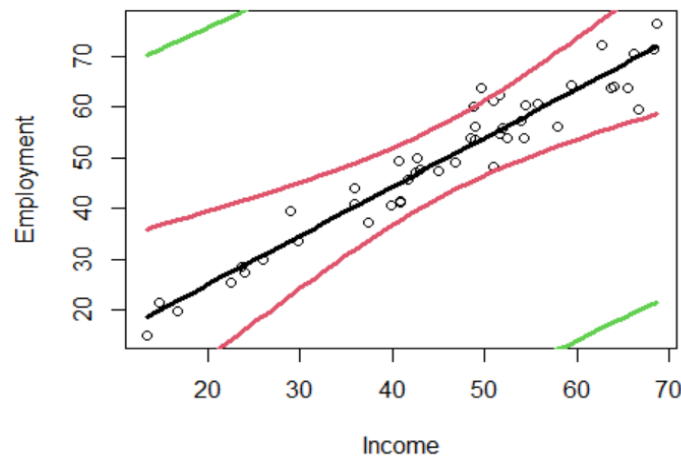


Figure 5. Plot of CI and PI of income and employment

3.4 The prediction of model

Then this paper will use backward selection in order to choose the best model between appropriate independent variables. Every time after finishing a regression, remove the independent variable which has the largest p-value among all the independent variables. After removing one independent variable, do the next regression and loop this step until each of the independent variables has a very small p-value.

After four times backward selection method, moving four different factors which are Education, Living, Housing, Health. As a result, the regression equation is:

$$\text{Income} = \beta_1 \text{Employment} + \beta_2 \text{Crime} + \beta_0 + \varepsilon \quad (4)$$

Then calculating the parameters in the equation, the final result is

$$\text{Income} = 0.84476 \text{Employment} + 0.14639 \text{Crime} - 3.2995 + \varepsilon \quad (5)$$

At this point, the research already leads to constructing a multiple linear regression model between Income and Crime and Employment. At last is the part of discussion. First and foremost, the study researches at how people's income changed in the after the coronavirus epidemic. Most previous studies may focus on employment and explore the change of people's income caused by the change of employment, but this study considers and analyzes the impact of various factors on income. But on the other hand, this study also contains some shortcomings. First of all, the final model of this study shows that the correlation between income and most factors in the data used is relatively weak, which is a little different from the conclusions of previous studies referred to. It is likely that the objects studied in this paper are different from those studied in other articles. In order to improve this, this study should collect more data related to income for analysis, so that it is possible to obtain a more accurate model than the research results of the revised paper. Secondly, there are deficiencies in the modeling methods used. This study show that establish a correct mathematical model and the using regression analysis to judge the significance is exceedingly important, which can make the research results more authentic and universal. This research only uses R studio to build a model between Income and different factors. However, this paper can also use other software to establish this model, which these can not only test the accuracy of the model, but can also compare with the model that using R studio to build, and at last choose the most reasonable result. Last but not least is the limitations of the model that established. This study only used data from the UK and did not use data from other countries or regions for analysis. As a result, the model may only be appropriate when predicting the incomes of people in the UK during the coronavirus epidemic. For different countries, the factors affecting income may be fundamentally different from those in the UK. Therefore, the model established by this study is only applicable to the prediction of the income of the British people under the epidemic wave.

4. Conclusion

The last part of the paper is the conclusion. The purpose of this paper is to explore whether there are factors that have a great impact on people's income in the UK under the influence of a wave of epidemic. First and foremost, the study shows that there is a strong correlation between employment and income, and the correlation coefficient reaches 0.96. In addition, log model and sqrt model of independent variables were used to draw Q-Q plot and residual plot respectively. It can be found that the two are not independent of each other, which further proves that income and employment have a strong correlation. Then the confidence interval and prediction interval are calculated, and it can be found that most points fall within the confidence interval and prediction interval. The next step is model building. Backward Selection method was adopted in this study. Firstly, all independent variables were included in the regression model, and after multiple model optimization, which leads to a multiple regression equation of income, employment and crime. Through this study, this paper can draw a conclusion: in a wave of epidemic, employment and crime have the greatest impact on the income of British people, while other factors, such as education and health, are not proved to have the same significant impact on income. The model can also be used to forecast the future, but it will be limited in the UK. The conclusions of this research can help enterprises to find a balance between employees' salary and employment to maintain revenue growth, and provide a variety of ideas for enterprises to reform the existing system in the wave of the epidemic. This paper also provides evidence that crime has a significant impact on income, providing more reasons to the importance about supervise crime.

References

- [1] Razai M S, Doerholt K, Ladhani S, et al. Coronavirus disease 2019 (covid-19): a guide for UK GPs [J]. Bmj, 2020, 368.

- [2] Johnson M T, Johnson E A, Webber L, et al. Mitigating social and economic sources of trauma: The need for universal basic income during the coronavirus pandemic [J]. *Psychological Trauma: Theory, Research, Practice, and Policy*, 2020, 12(S1): S191.
- [3] Adams-Prassl A, Boneva T, Golin M, et al. Inequality in the impact of the coronavirus shock: new survey evidence for the UK [J]. 2020.
- [4] Kholodilin K A. Housing policies worldwide during coronavirus crisis: Challenges and solutions [J]. 2020.
- [5] Arnold L. Understanding and promoting autonomy in UK online higher education [J]. *International Journal of Instructional Technology & Distance Learning*, 2006, 3(7): 33-46.
- [6] Fazey D and Fazey J 'The potential for autonomy in learning'. *Studies in Higher Education*, 2001, Vol. 26 no 3, 345-361.
- [7] Dickinson L. Autonomy and motivation. *System*, 1995, 23 (2), 165-174.
- [8] Deaton A. Health, income, and inequality [J]. *National Bureau of Economic Research Reporter: Research Summary*. Retrieved August, 2003.
- [9] Zhang J, Lu X, Jin Y, et al. Hospitals' responsibility in response to the threat of infectious disease outbreak in the context of the coronavirus disease 2019 (COVID-19) pandemic: Implications for low-and middle-income countries [J]. *Global Health Journal*, 2020, 4(4): 113-117.
- [10] Patterson E B. Poverty, income inequality, and community crime rates [J]. *Criminology*, 1991, 29(4): 755-776.
- [11] Kim B, Seo C, Hong Y O. A systematic review and meta-analysis of income inequality and crime in Europe: Do places matter? [J]. *European Journal on Criminal Policy and Research*, 2020: 1-24.