

Impact of Coronavirus Pandemic on US Citizens Income

Alina Khindanova, Anvit Sachdev, Shreya
Kamble

Abstract

This study investigates the impact of the COVID-19 pandemic on total personal income, using data from the Current Population Survey from 2017-2020. We constructed a model to predict individual income based on data from 2017-2019, using three different machine learning algorithms. The results showed that the random forest model had the best performance, with the smallest root mean square error. Using this model, we predicted personal income for 2020 and compared it to actual income values. The findings suggest a significant difference between income and fitted values, indicating a positive impact of the pandemic on personal income.

In summary, this study sheds light on the complex relationship between the COVID-19 pandemic and personal income, highlighting the potential for unexpected positive effects during times of crisis. The use of advanced machine learning techniques provides a powerful tool for analyzing and understanding this impact and may have important implications for policymakers and researchers working in related fields.

Introduction

On January 21, 2020, the Centers for Disease Control and Prevention (CDC) announced the first COVID-19 case in the USA. After that, the pandemic was spreading rapidly, and the COVID Data Tracker of CDC shows that the number of new coronavirus cases reached its peak two years later after the beginning of the pandemic – on January 19, 2020.

The COVID-19 had a huge effect on population health: it was recorded as the underlying or contributing cause of 378,000 deaths nationwide (Ahmad et al. 2021), however the effect was different for different subgroups of people (Alsan et al., 2021). Since the health status is considered as one of the predictors of income, the coronavirus epidemic could negatively affect the income. What is more, the pandemic ruined a considerable part of business operations which led to a significant increase in unemployment rate (Couch et al., 2020, Fairlie et al., 2020) – that could also decrease personal income.

On the other hand, the US government made a big contribution towards human welfare: there were direct transfers to US citizens (three rounds of Economics Impact Payments), increase in unemployment benefits (extension of employment assistance, some federal taxes were waived), less strict requirements in the eligibility of public insurance (expansion of Medicaid and Medicare), and also support of local businesses (tax benefits, financial support).

Therefore, the effect of the COVID-19 pandemic on personal income is not straightforward. This paper studies the effect of the pandemic on total personal income. Since the epidemic started quite recently, there is not much literature on this topic. Previous studies estimate the effect on particular characteristics (Esobi et al., 2021), or certain group of people (for example, van Kooten and Schmitz, 2022, Nguyen and Schmitz, 2023). We contribute to the existing literature by considering the whole US adult population and using machine learning techniques to estimate the effect of the pandemic on personal income.

We use information from the Current Population survey (CPS) from 2017-2020. We restrict our sample to individuals who are older 20 years. We split this data to training set (2017-2018) to fit the models and testing set (2019) to check their performance. We use three different models to predict individual income: K-nearest-neighbors (KNN), gradient-boosted tree and random forest. Our estimated effect is the difference between actual and predicted outcomes for the sample of 2020.

Methods

We use information from the Current Population survey (CPS) from 2017-2020. Our outcome variable is individual's total pre-tax personal income or losses. The set of features that we used to fit our models is:

- Age
- Sex – whether a person is male/female
- Marital status
- Relationship to household head – we encoded initially categorical variable as dummy variable whether a person is head of a household or not
- Race
- State
- Number of children in the family
- Citizenship status
- Employment status - we encoded initially categorical variable as dummy variable whether a person is working or not
- Occupation
- Hours usually worked per week
- Educational attainment recode
- Self-reported health

We restrict our dataset to individuals older 20 years old. Additionally, we exclude individuals whose hours usually worked per week vary.

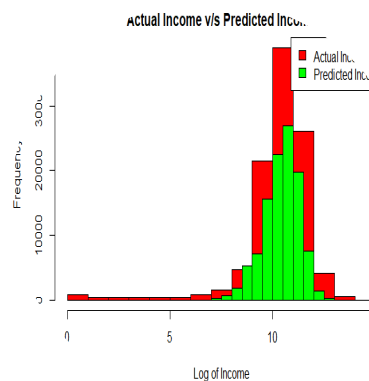
We use three different regression models to predict individual income: K-nearest-neighbors (KNN), random forest and gradient-boosted tree. In order to fit the model, we use data for the years 2017-2018. We have used cross-validation in conjunction with hyperparameter tuning to fit the models. The test set comprises of the data points from the year 2019. We have used root mean squared error as the evaluation metric. Additionally, we have also tested the performance of our models for the year 2020. For the model that best fits the data, we have used t-test to test whether the mean of difference between actual and predicted values is statistically different for the years 2019 and 2020. This is done to check whether the impact of covid on an individual's income was statistically significant.

Results

The below table shows the root mean squared error of the three models for the years 2019 and 2020:-

Model	RMSE (2019)	RMSE (2020)
KNN	59324.26	64277.83
Random Forest	58304.54	63211.05
Gradient-Boosted Tree	59354.69	64151.26

We can see from the above table that random forest is performing the best. Furthermore, we can see that for each model, the RMSE of the year 2020 is more than the RMSE of 2019. This indicates that the models fitted on 2017 and 2018 data are not able to explain the incomes of the individuals in 2020, implying that incomes in 2020 have been impacted. To gather further evidence, we used t-test to test whether the mean of difference between actual and predicted values is statistically different for the years 2019 and 2020. We find that the mean of the difference between actual and predicted values for the years 2019 and 2020 is 1223.556 and 4801.022, respectively. The p-value between them is less than $2.2e-16$ implying that the difference between the two values is statistically different. The below figure shows the histogram of log of income for the years 2020:-



The above figure shows the relationship between actual income and predicted income, with the income on the x-axis and frequency on the y-axis. The plot of actual income value is above the predicted income values in 2020. We can see that the pandemic has affected different different parts of the population in different ways. The plot indicates that a lot of people's income shifted to the middle-income group during 2020. We can also see from the plot people's income shifted to a lower category than the pre-covid years.

Conclusion

Given that the data is related to income, the results suggest that there was a major impact of COVID-19 on income between 2019 and 2020. Specifically, the mean difference in income between the two years was significantly lower in 2020 compared to 2019. This could be due to variety of factors, such as job loss, reduced work hours, or lower wages resulting from the pandemic.

The histogram shows the relationship between actual income and predicted income, with the income on the x-axis and frequency on the y-axis. The plot of actual income value is above the predicted income values in 2020. This suggests that there may have been a negative impact of the COVID-19 pandemic on people's income. The pandemic has affected different age groups and different parts of the population in different ways.

The plot indicates that a lot of people's income shifted to the middle-income group during 2020, this could be due to several factors related to the COVID-19 pandemic. For example, middle aged workers have been particularly hard hit by the pandemic, as they are more likely to work in sectors that have been heavily impacted. This could have resulted in many people experiencing a decrease in income that caused them to move from a higher income group to a middle-income group. We can also see from the plot people's income shifted to a lower category than the pre-covid years. This could be due to the same factors as above. For example, job loss, reduced work hours, or lower wages resulting from the pandemic could have caused many people to experience a decrease in income that caused them to move from a middle-income group to a low-income group.

References

- Ahmad, Farida B., Jodi A. Cisewski, Arialdi Miniño, and Robert N. Anderson. 2021. "Provisional Mortality Data — United States, 2020." *Morbidity and Mortality Weekly Report* 70 (14): 519–22.
- Alsan, Marcella, Amitabh Chandra, and Kosali Simon. "The great unequalizer: initial health effects of COVID-19 in the United States." *Journal of Economic Perspectives* 35.3 (2021): 25-46.
- Couch, Kenneth A., Robert W. Fairlie, and Huanan Xu. "Early evidence of the impacts of COVID-19 on minority unemployment." *Journal of public economics* 192 (2020): 104287.
- Fairlie, Robert W., Kenneth Couch, and Huanan Xu. The impacts of COVID-19 on minority unemployment: First evidence from April 2020 CPS microdata. No. w27246. National Bureau of Economic Research, 2020.
- Esobi, I. C., et al. "Food insecurity, social vulnerability, and the impact of COVID-19 on population dependent on public assistance/SNAP: A case study of South Carolina, USA." *Journal of Food Security* 9.1 (2021): 8-18.
- van Kooten, G. Cornelis, and Andrew Schmitz. "COVID-19 impacts on US lumber markets." *Forest policy and economics* 135 (2022): 102665.
- Nguyen, Ly, and Andrew Schmitz. "The welfare impacts of Covid-19 on the US salmon sector." *Applied Economics* 55.22 (2023): 2579-2595.