# **Impact of Unemployment on the Household Debt by County**

##### Lenin Kamma

##### Bellevue University

##### 1000 Galvin Rd S, Bellevue, NE 68005

# **Executive Summary**

# We are facing a global health crisis unlike any other crisis in the history of the world. This health crisis is has also turned into a financial and social crisis1. Unemployment is on the rise and there is an increase in non-performing loans and debts. As part of this study, predictive analysis is applied to detect the counties across the nation where unemployment and debt are rising fast. This study is also focused on the correlation between unemployment and debt.

# It has been found that unemployment is rising faster than expected. The top 100 counties with the highest unemployment in 2019 have almost doubled their numbers in 2020. The household debt has also been on the rise since 2019’s first quarter to the last quarter. An interesting fact confirmed by the study is that there is no direct correlation between unemployment and the debt. The counties where the debt is at the highest, unemployment is not at the highest. So, other factors not in the scope of this study are impacting the household debt. One important thing discovered in this study is that unemployment is still less than two percent in 450 counties across the nation. While Michigan tops the counties with the highest unemployment, Kentucky, Alabama, and Georgia top the list of highest debt. There is a chance the household debt in 2020 could rise as a result of unemployment although no direct correlation is observed. The magnitude of the pandemic’s impact is unlike anything we have experienced in our lifetimes.

# **Introduction**

The cost of unemployment to the individual is well known. When a person loses their job, there is often an immediate impact on that person’s standard of living. That is more evident than ever, due to the Covid-19 pandemic, which left 10 million Americans jobless in its first two weeks2. Financial companies especially mortgage and credit card companies are severely impacted by unemployment as payment defaults are expected to increase over the next several months. Analytical models are being developed by various companies to predict the future performance of borrowers. However, these models operate at the macro level by using national unemployment statistics. In this study, I followed a different approach by taking unemployment and household debt data at the county level. The models were developed to study unemployment at each county and predict which county is going to have severe unemployment in the future. This data was then compared with the household debt data by county to see if there was any correlation between debt and unemployment.

# **Methods**

## Importing the Data

This study has used unemployment and debt data by county. The datasets were imported as CSV files from Bureau of Labor statistics and Federal Reserve websites.

The unemployment file has the following variables

* Area code State code County code County name
* Labor force Employed Unemployed Unemployment rate

The debt file has the following variables

* Year Quarter FIPS code Low debt
* High debt

## Data Validation

Data validation was done by comparing basic statistics with published results on the bureau of labor website. The data check process has been used to remove duplicate rows per county. The imported datasets had some errors and invalid values in county columns. To correct the data, cleanup was done to remove NA or null values before creating the dataframe. Summary statistics were also cleaned up as part of data validation.

## Data Transformation:

As FIPS code (county code) is a key variable for exploratory analysis, it was created by concatenating both the state code and county code columns. The following data transformation processes were carried out as part of this study.

* + Removed leading zeros in numeric columns
  + Filled missing unemployment-high values with unemployment-low values
  + The variable period was split into month and year and removed all commas in numbers
  + Dropped data rows where unemployment and debt values were not present
  + The dropped rows were less than 0.1 percent of the whole dataset
  + Converted string values to numerical values for comparison purpose
  + Converted month and year values to match with quarterly values in debt file
  + Removed extra spaces in unemployment rate values
  + Final data cleanup was completed by dropping columns which were not required for this study
  + Calculated mean value of unemployment rate for each quarter
  + Used county code, year and quarter to merge debt data with unemployment data

# Model development:

As part of model development many scatter plots and bar charts were plotted to study different variables and their characteristics. The Pearson coefficient and Spearman coefficients were measured to find the correlation between the household debt and unemployment to quantify the strength of the relationships. Linear regression was also carried out to predict the debt rate from unemployment.

* Scatter plots
  + - * + Unemployment rate by county
        + Household debt by county
        + Unemployment rate of top 100 counties vs Debt rate
        + Debt Rate of top 100 counties vs Unemployment rate
        + Unemployment rate vs Debt rate of all counties
* Bar Charts
  + - * + County vs Unemployment rate for 2019 and 2020
        + County vs Household Debt for 2019
        + Top 100 counties with the highest unemployment in 2019
        + Top 100 counties with the highest unemployment in 2020
        + Top 100 counties with the highest debt in 2019
* Correlation analysis between unemployment and debt rate of all counties
* Regression analysis to find the debt rate based on unemployment
* Split data into training (40%) and test (60%) datasets to complete the regression
* Measured Pearson and Spearman coefficients
* Measured accuracy of the linear regression model

# Results

* In 2019, Kusilvak Census Area (AK) has the highest unemployment rate of 24% followed by imperial county, CA with 21%
* More than 100 counties have unemployment rate of greater than 12% in 2019
* The unemployment rate is doubled from 2019 to 2020 in100 counties with the highest unemployment
* The highest debt rate in top 100 counties found to be at 3.46
* The top 100 counties with unemployment do not carry much debt
* Haines Borough AK has the highest unemployment rate (12.03) and the highest debt rate (3.46)
* The highest debt rate is found in many counties across Kentucky, Indiana, Illinois, Georgia and Colorado
* Loving county, TX and Kiowa county, CO has the lowest unemployment rate of 0.60
* 450 counties in US have unemployment rate less than 2%
* The correlation coefficient value was found to be -0.02800 from correlation matrix
* The Pearson correlation coefficient for unemployment was at -0.028009257637667588
* The Spearman correlation coefficient was found to be 0.016740407712785045
* The accuracy of the regression model was observed to be 0.012

# Discussion/Conclusion

* It was observed that the unemployment rate for the top 100 counties in 2019 has almost doubled in 2020, which is a cause for concern. For some counties, the debt rate has also slightly increased from 1st quarter to 4th quarter in 2019
* While Michigan has many counties with the high unemployment rate, Kentucky, Georgia, and Alabama have counties with a very high debt rate. This has shown that the debt rate and unemployment rates are not uniform across the country.
* Around 450 counties have an unemployment rate of 2 percent. However, there are more than 25 counties with a higher unemployment rate of more than 25%.
* The low value of correlation coefficient (-0.028) and Pearson coefficients (-0.028) indicate that no correlation exists between the unemployment rate and the debt rate. This is also evident from models where the counties with the highest unemployment rate have very low debt rate.
* The low value of regression model accuracy also indicates that it is not possible to predict the debt rate based on unemployment as other factors have an impact on the household debt, which are out of scope for this study.
* There are many counties with a very low unemployment rate (less than 1%) but with a very high debt rate (>30%)

# Limitations:

This study is carried out with the unemployment data available from 2019 to May 2020 and household debt data available only for 2019. The household debt data for 2020 was not available at this time. There are other factors that have an impact on household debt data which are not included in this study.

# Acknowledgements

I would like to express my sincere thanks to Professor Fadi Alsaleem for his continuous support during the course. His guidance and feedback every week helped me complete this project. Along with my professor, I would like to thank my classmates for their continued support, review, and feedback needed ­­­­for the successful completion of this project.

# References

1 https://www.un.org/development/desa/dspd/everyone-included-covid-19.html

2 https://www.investopedia.com/financial-edge/0811/the-cost-of-unemployment-to-the-economy.aspx

3 https://www.bls.gov/web/metro/laucntycur14.txt

### 4 County-Level Debt-to-Income Ratio, 1999 – 2019, https://www.federalreserve.gov/releases/z1/dataviz/household\_debt/county/map/#state:all;year:2019

(Images are given below)

# Images:









