Database design to see the impacts of COVID-19 on the US Labor Market

Final Project Paper

EPPS 6354

Prattasha Nawar Islam

Spring 2023

Abstract

This paper is based on a project of designing and creating an original database application. The database application is to see the effects of the COVID-19 recession on the labor market indicators of the US. The analysis shows that due to different COVID restrictive policies of school and workplace closing, the unemployment rates of men and women increased significantly. However, there was a noticeable difference in the unemployment rates of men and women. Women's unemployment rate was higher than men possibly because of the working mother giving up their jobs to take care of their children as the schools and day-care centers were closed during that time.

Introduction

This paper discusses the design of the database for the use of observing the impacts of COVID-19 policies on the US labor market dynamics. To be specific, the paper looks at how the COVID-19 recessionary period affected the labor market indicators in the US economy. COVID-19, even though it started in the year 2020, it still seems to have brought some permanent changes in the labor market even today. From working remotely or in a hybrid modality to having most work done online, COVID did bring in some drastic changes in a lot of industries around the world.

An interesting aspect of the labor market changes is that it affected men and women workers differently. According to the researchers in Federal Reserve Economic Data (FRED), recessionary periods in general have always had a larger impact on men's employment in comparison to women. And they introduced two new terms for this difference in impact. Hence, it was always a "he-cession" and then followed by a recovery period of "he-covery" in the previous case of recessions. However, FRED now claims that looking at the statistics after the COVID-19 recession, it is seen to be trending towards a "she-cession" instead. This means they are claiming that women's employment got hit harder than men's employment after COVID.

The creation of this database will be useful for researchers to use as a tool for studying the impacts of the pandemic on the labor market. The database can be used for the purpose of any research, especially for economic and social data research. The database would be even more useful if it is paired up with relevant econometric model for analyzing the relationship accurately.

Data

The data used to design the database is collected from the FRED online database. The FRED data series proposed to be used are all time series data series. The data series initially proposed and shortlisted to explain the study better are as follows,

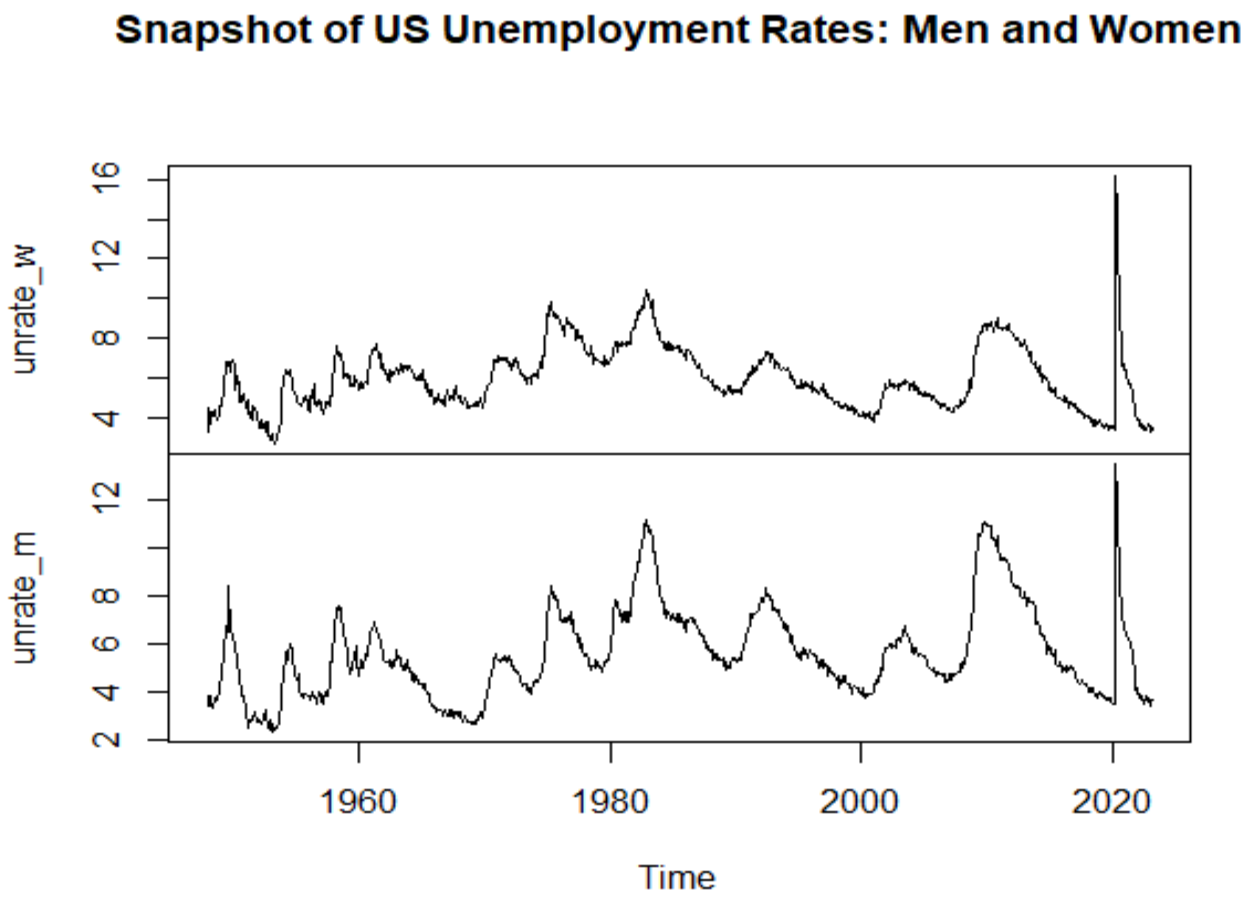
- Unemployment Rate – Men
- Unemployment Rate – Women
- Labor Force Participation Rate – Men
- Labor Force Participation Rate – Women
- Employment-Population Ratio – Men
- Employment-Population Ratio – Women

The above data series are all used in the percent unit in monthly frequency. They are all seasonally adjusted by FRED. The data started in January 1948 and was updated till March 2023 at the time of collecting the data.

Two data series shortlisted from the Our World in Data platform are:

- COVID - 19: School and Workplace Closures
- COVID – 19: Income Support and Debt Relief

The two data series are under the policy responses to the coronavirus pandemic section of the Our World in Data platform. The time span of the data is between January 2020 till December 2022. The data was converted to monthly data by selecting the observations at the first of each month.



The above chart shows a glimpse of the full sample data of unemployment rates of men (unrate_m) and women (unrate_w). Because the project investigates the COVID years (2020 and beyond), a strong spike can be seen in the year 2020. To study it closely and for the simplification of the project and to make it more feasible during the time of the creation of the database application, the above data was sampled and shortened to a smaller time span just to observe the immediate effects of COVID policies. The data time span was shortened between March 2020 till March 2021. Moreover, only two of the data series from each data source were used to create the database application. The data series selected and sampled for the database application are as follows,

- Unemployment Rate – Men

- Unemployment Rate – Women
- COVID - 19: School Closures
- COVID -19 Workplace Closures

In the above data series of COVID-19 school and workplace closures, the data is coded as below,

School closures:

0 – No measures

1 – Recommended closing

2 - Require closing (only some levels or categories, e.g., just high school, or just public schools)

3 - Require closing all levels

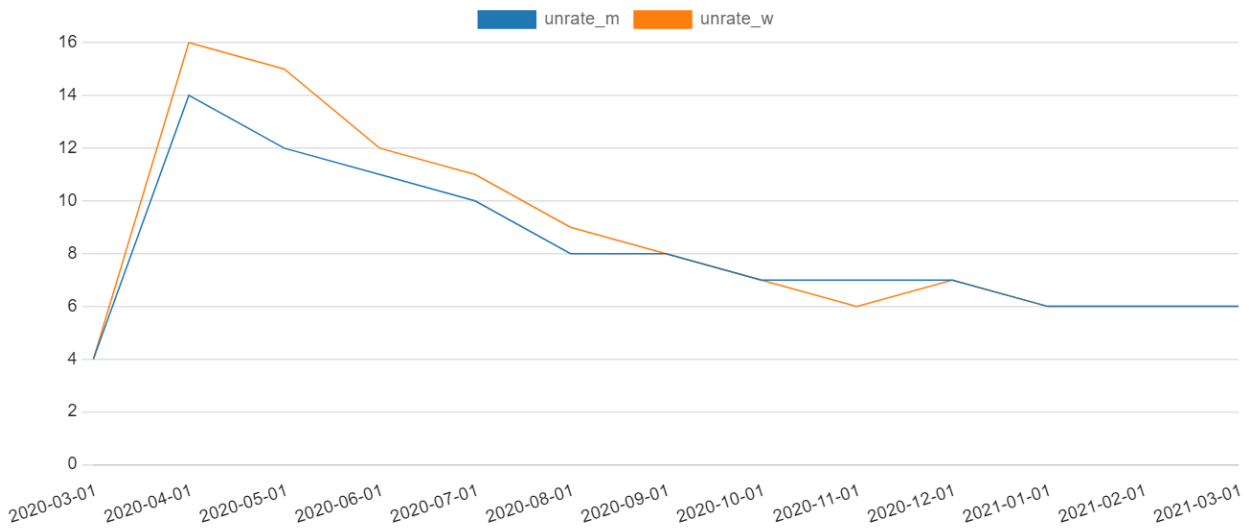
Workplace closures:

0 – No measures

1 – Recommended closing (or work from home)

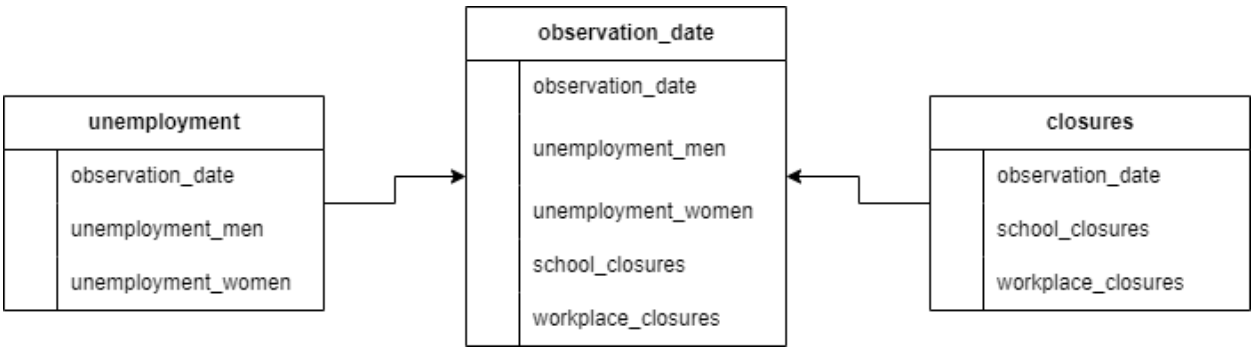
2 - Require closing or work from home (for some sectors or categories of workers)

3 - Require closing all levels (for all but essential workers)



The spike in 2020 that could be seen in the full-sample chart is now zoomed in as a subsample of the data as in the graph above.

A rough schema is shown below.



Methodology

As the end goal of this database design is to create an R Shiny application of the database, the process started with inputting the data in the pgadmin 4 application using PostgreSQL. First, two tables were created, named “unemployment” for the unemployment data and “closures” for the

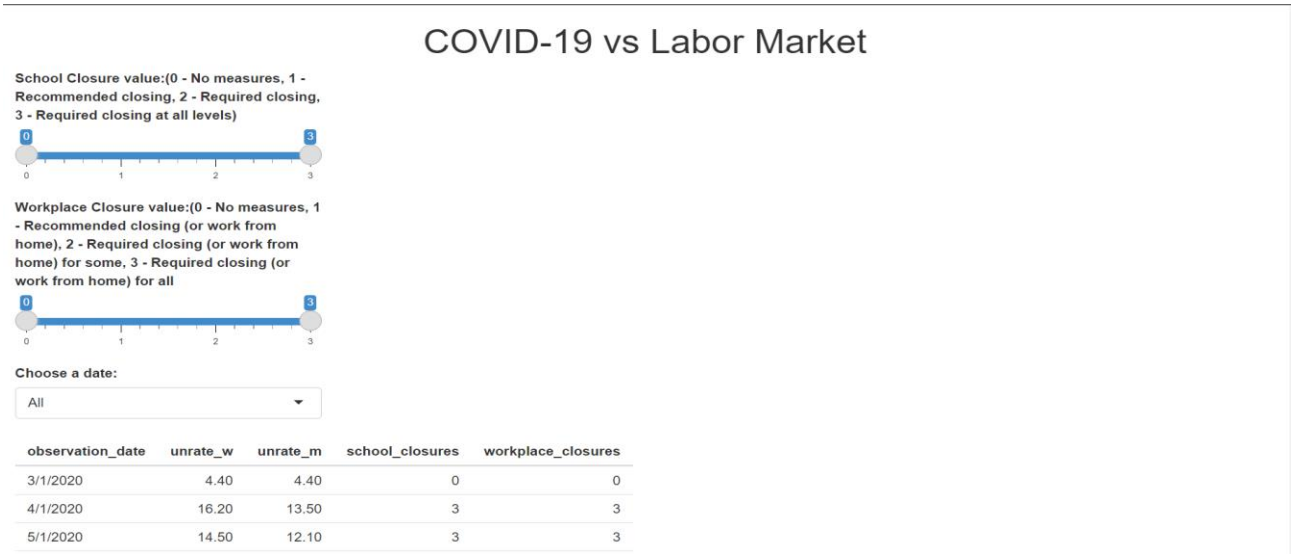
school and workplace closures data. Next, data was inserted into the tables into different columns with “observation_date” being the common column for both the columns. After the data was inserted into the tables, both the tables were pulled out together joined by the date column using Data Definition Language (DDL) and Data Manipulation Language (DML) statements. A glimpse of the final SQL file is shown below.

```
CREATE TABLE unemployment (
    observation_date    DATE,
    unrate_w            NUMERIC(4,0),
    unrate_m            NUMERIC(4,0),
    lfpr_w              NUMERIC(4,0),
    lfpr_m              NUMERIC(4,0)
);
CREATE TABLE closures (
    observation_date DATE,
    school_closures NUMERIC(4,0),
    workplace_closures NUMERIC(4,0)
);
INSERT INTO unemployment(observation_date, unrate_w, unrate_m, lfpr_w, lfpr_m) VALUES ('3/1/2020',4.4,4.4, 57.1, 68.5),
('4/1/2020',16.2,13.5, 54.6, 66.1),
('5/1/2020',14.5,12.1, 55.2, 66.7),
('6/1/2020',11.6,10.5, 56, 67.3),
('7/1/2020',10.7,9.7, 56.1, 67.2),
```

Then, in a new R studio environment, all the necessary packages were installed, and the R Shiny app was started. The Shiny app has its ui and server inputs and outputs connected to the database created in pgadmin 4. The database was then converted to an SQLite format using the DB browser for SQLite and then imported into R studio. Finally, the Shiny app was deployed and published in R studio.

Finding and Analysis

Even though the application (https://prattashaislam.shinyapps.io/FinalProject/) did not deploy on the web properly due to some technical issues, it rendered well locally. The output of the database design is as shown in the following images.



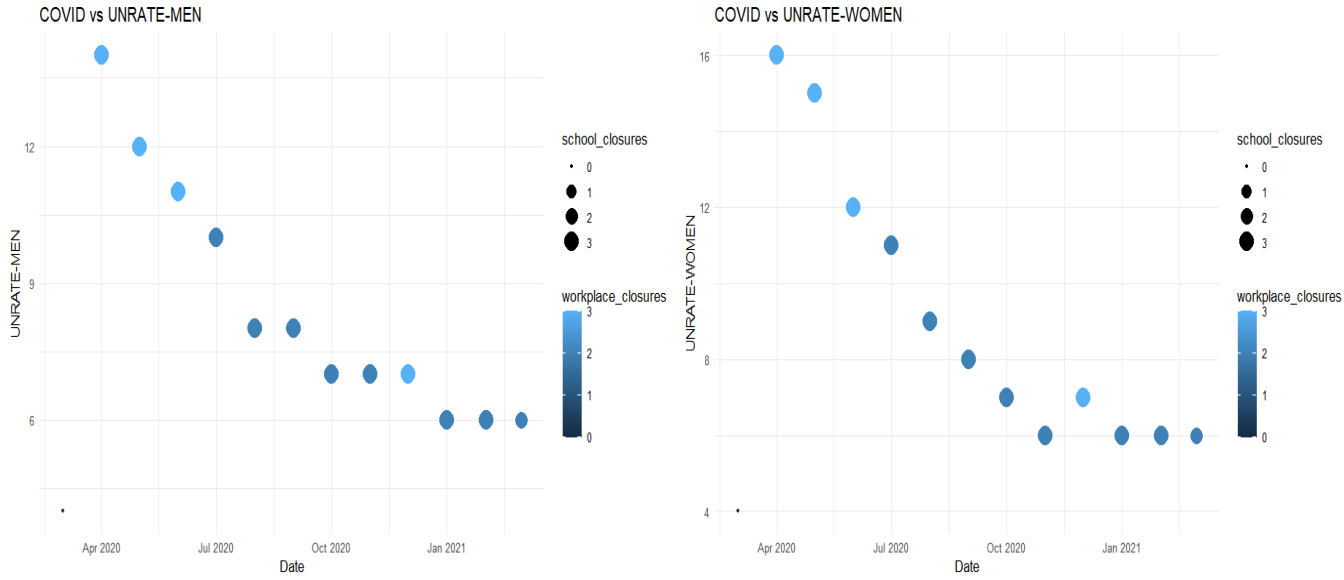
The database starts with a title, “COVID-19 vs Labor Market” which explains that this database can be used to investigate the cases of the pandemic and its effects on the labor market. The first interactive tool on the database is a slider input for the category of selecting “school closures” values between 0 and 3 as discussed above in the data section. Similarly, another slider input for the category of selecting “workplace closures” values is placed right below. The slider inputs are followed by a select input of a list of the dates available including “All” option, where all the dates in the database can be observed at once.

Choose a date:

All

observation_date	unrate_w	unrate_m	school_closures	workplace_closures
3/1/2020	4.40	4.40	0	0
4/1/2020	16.20	13.50	3	3
5/1/2020	14.50	12.10	3	3
6/1/2020	11.60	10.50	3	3
7/1/2020	10.70	9.70	3	2
8/1/2020	8.60	8.20	3	2
9/1/2020	8.10	7.70	3	2
10/1/2020	6.70	7.00	3	2
11/1/2020	6.40	6.90	3	2
12/1/2020	6.70	6.70	3	3
1/1/2021	6.30	6.40	3	2
2/1/2021	6.10	6.40	3	2
3/1/2021	5.90	6.20	2	2

The image above shows the output table rendered in the database for all observations. In March 2020, there were no restrictions in the school and workplace closure policies and hence the unemployment rates for men and women were the same value of 4.40 percent. Whereas, when COVID first hit the economy, the restrictions started coming with some school and workplace closing, this led to unemployment rates of men and women to spike up sharply. However, the differences in the unemployment rates of men and women can be seen from the above table. Women's unemployment was at 16.20 percent in April 2020, whereas men's unemployment rate was at 13.50 percent. This difference in unemployment rates were seen until September 2020, after which both the rates came back to equal levels again and went back to almost the usual levels by the start of 2021, where women's unemployment rate seemed to be slightly lower than men.



The last part of the database design contains the data visualization part where the R packages, ggplot2 and gridExtra are used to show two separate scatter plots to show and compare the unemployment rates of men and women against time. It is observed that the data is trending downwards as the graph time span goes from the start of the pandemic to a year after the pandemic. The school closure values are assigned to be the size of the data points and the workplace closure values are assigned to be the color shades of the data points. So, if noticed carefully, the data points in March 2020 for both the series are tiny little dots which indicate that the school closure values were at 0 so the size is the smallest and the workplace closure values were at 0 as well so their color is the darkest shade of blue.

FRED explains one possible reason for this difference in men and women's unemployment rates when the pandemic initially hit, is the closing of schools and day care centers. So, this situation probably put most of the childcare responsibilities disproportionately on mothers, or in other words, the female workers in the economy. Hence, there were more women who became unemployed initially.

Conclusion

In conclusion, this project is about designing and creating an original database application to be used for the purpose of research. A R Shiny application was created with the database design done in R studio. The database is to analyze the impacts of the COVID-19 pandemic policies on the labor market indicators like unemployment rates of men and women. In the analysis, it is observed that during the COVID period, more women lost their jobs compared to men due to the policy measures of school and workplace closing. However, under separate research of mine, I conducted two autoregressive models of order 1 (AR1) to see the effects of a post-covid dummy variable where dates before March 2020 are set equal to 0 and dates after March 2020 are set equal to 1, on the unemployment rates of men and women. The result of the regression gives no statistically significant coefficients for the post-covid dummy variable. Hence, this indicates that even though there was an impact and differences seen in the unemployment rates due to the pandemic, the effects were not large enough to create any structural changes in the time series pattern. For this reason, the spike in the graph seen in 2020 came down to normal level very rapidly. In future, the project will have solutions to any technical issue with the application building and design the database to be more interactive and customizable. Also, replicating the project with more data observations and relevant series from different sources would make the database even more accurate and useful.

References

Edouard Mathieu, Hannah Ritchie, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Joe Hasell, Bobbie Macdonald, Saloni Dattani, Diana Beltekian, Esteban Ortiz-Ospina and Max Roser (2020) - "Coronavirus Pandemic (COVID-19)". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/coronavirus'

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

R packages used: Shiny, RPostgres, Pool, DBI, ggplot2, gridExtra

Silvio Contessi and Li Li, "From "Man-Cession" to "He-Covery": Same Old, Same Old," *Economic Synopses*, No. 3, 2013. https://doi.org/10.20955/es.2013.3

The covid-19-induced "she-cession". FRED Blog. (n.d.). Retrieved from https://fredblog.stlouisfed.org/2020/11/the-covid-19-induced-she cession/?utm_source=series_page&utm_medium=related_content&utm_term=related_resources&utm_campaign=fredblog

U.S. Bureau of Labor Statistics, Unemployment Rate - Women [LNS14000002], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/LNS14000002, May 7, 2023.

U.S. Bureau of Labor Statistics, Unemployment Rate - Men [LNS14000001], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/LNS14000001, May 7, 2023.