Final Project

Shanshan Liu, Michael Vierela, Jeremy Sternlicht, Charlie Gartenberg

## Objective

The following analysis seeks to understand the effects of the COVID-19 pandemic on the retail industry, in terms of unemployment, and this impact in relation to the broader economy.

## **Data Source**

We retrieved employment data from the Current Population Survey (CPS) in the IPUMS database. We included data on employment status to determine the probability of unemployment, gende r, race, family income, education level, and other industries to evaluate if unemployment was impacted by any other factors associated with the population and industry. We restricted this data from one year before to one year after the pandemic (March 2019 - February 2021).

## **Assumptions**

The reason why we restricted the data for a year before and after the pandemic was because we didn’t want the overall effect of the pandemic to be swayed by too much by its initial impact or diluted by the post-pandemic recovery. We also used the month of each year as a categorical variable to demonstrate how the labor market recovered over time. Data prior to March 2020 was considered “before pandemic” and data from March 2020 onward was considered “after pandemic” because pandemic-related restrictions and closures started in the United States during March 2020.

On an industry level, we also decided to separate arts, entertainment, and recreation, and accommodation and food services from other services because of many pandemic-related restrictions that specifically focused on hospitality and leisure venues where people congregate like restaurants, bars, and theaters.

## Libraries

In the following code chunk, load all the libraries we will need:

#load packages  
  
library(rio)

Warning: package 'rio' was built under R version 4.3.2

library(stringr)  
library(lubridate)

Warning: package 'lubridate' was built under R version 4.3.3

Attaching package: 'lubridate'

The following objects are masked from 'package:base':  
  
 date, intersect, setdiff, union

library(dplyr)

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':  
  
 filter, lag

The following objects are masked from 'package:base':  
  
 intersect, setdiff, setequal, union

library(vtable)

Warning: package 'vtable' was built under R version 4.3.2

Loading required package: kableExtra

Warning: package 'kableExtra' was built under R version 4.3.2

Attaching package: 'kableExtra'

The following object is masked from 'package:dplyr':  
  
 group\_rows

library(fixest)

Warning: package 'fixest' was built under R version 4.3.2

library(ggplot2)

## Load Data

# data from ipumsr  
  
if (!require("ipumsr")) stop("Reading IPUMS data into R requires the ipumsr package. It can be installed using the following command: install.packages('ipumsr')")

Loading required package: ipumsr

Warning: package 'ipumsr' was built under R version 4.3.3

ddi <- read\_ipums\_ddi("cps\_00001.xml")  
data <- read\_ipums\_micro(ddi)

Use of data from IPUMS CPS is subject to conditions including that users should cite the data appropriately. Use command `ipums\_conditions()` for more details.

# read industry name data   
  
indnames <- rio::import('indnames.csv')

## **Merge Data**

# Merge two data frames based on the columns "IND" and "ind"  
  
data1 <- merge(data, indnames, by.x = "IND", by.y = "ind", all.x = TRUE)

## Create Dummy Variable & Clean Data

# Create industry dummy variables  
  
# Create Service industry dummy = 1, if it is the following industry:  
# Construction.  
# Information.  
# Other Services, Except Public Administration  
# Finance and Insurance, and Real Estate and Rental and Leasing  
# Transportation and Warehousing, and Utilities  
  
# Create hospitality industry dummy = 1, if it is the following industry:  
# Arts, Entertainment, and Recreation, and Accommodation and Food Services.  
  
# Create MS (Management & Science) dummy =1, if it's the following industry:   
# Professional,Scientific, and Management, and Administrative and Waste Management Services.   
# Public Administration  
  
data2 <- data1 %>%  
 mutate(  
 agriculture = ifelse(IND >= 170 & IND <= 490, 1, 0),  
 manufacturing = ifelse(IND >= 1070 & IND <= 3990, 1, 0),  
 retail = ifelse(IND >= 4670 & IND <= 5790, 1, 0),  
 wholesale = ifelse(IND >= 4070 & IND <= 4590, 1, 0),  
 military = ifelse(IND >= 9670 & IND <= 9890, 1, 0),  
 service = ifelse((IND >= 6070 & IND <= 6390) | (IND >= 570 & IND <= 690) | (IND >= 6470 & IND <= 6780) | (IND >= 8770 & IND <= 9290) | (IND >= 7860 & IND <= 8470) | (IND >= 6870 & IND <= 7190) | (IND ==770), 1, 0),  
 ms = ifelse((IND >= 7270 & IND <= 7790) | (IND >= 9370 & IND <= 9590), 1, 0),  
 hospitality = ifelse(IND >= 8560 & IND <= 8690, 1, 0))  
  
# Create unemployment dummy variable, = 1 if Unemployed experienced work or Unempolyed new work  
  
data3 <- data2 %>%  
 mutate(unemployment = ifelse(EMPSTAT %in% c(21, 22, 20), 1, 0))  
  
# Create Race dummy variable and filter out blank  
data4 <- data3 %>%  
 mutate(  
 white = ifelse(RACE == 100, 1, 0),  
 black = ifelse(RACE == 200, 1, 0),  
 aapi = ifelse(RACE %in% c(650, 651, 652), 1, 0),  
 indigenous = ifelse(RACE == 300, 1, 0),  
 multiracial = ifelse(RACE >= 801 & RACE <= 830, 1, 0),  
 other = ifelse(RACE == 700, 1, 0)) %>%  
 filter(RACE != 999)  
  
# Keep ages of research subject within the range of 18 to 60 and filter out NIU and NILF employment status  
  
data5 <- data4 %>%  
 filter(AGE >= 18 & AGE <= 60) %>%  
 filter(EMPSTAT %in% c(1, 10, 12, 20, 21, 22))  
  
# Create Gender dummy variable  
  
data6 <- data5 %>%  
 mutate(  
 male = ifelse(SEX == 1, 1, 0),  
 female = ifelse(SEX == 2, 1, 0))  
  
# Create Education level dummy variable  
# under\_hs dummy variable = 1 if highest level of education is less than a high school diploma  
# hs\_diploma dummy = 1 if highest level of education is a high school diploma or equivalent   
# some\_college = 1 if highest level of education includes some college  
# associate = 1 if highest level of education is an associate's degree  
# bachelor dummy variable = 1 if highest level of education is a bachelor's degree  
# graduate dummy variable = 1 if highest level of education is higher than a bachelor's degree (Master, Professional school degree, PHD)  
# filter out NIU values  
  
data7 <- data6 %>%  
 mutate(  
 under\_hs = ifelse(EDUC >= 2 & EDUC <= 72, 1,0),  
 hs\_diploma = ifelse(EDUC == 73, 1, 0),  
 some\_college = ifelse(EDUC %in% c(80, 81, 90, 100, 110, 120, 121, 122), 1, 0),  
 associate = ifelse(EDUC %in% c(91, 92), 1, 0),  
 bachelor = ifelse(EDUC == 111, 1, 0),  
 graduate = ifelse(EDUC %in% c(123, 124, 125), 1, 0)) %>%  
 filter(!(EDUC %in% c(999, 0, 1)))  
  
# Create time dummy variable =1, if date after covid19 Mar.2020  
  
data8 <- data7 %>%  
 mutate(  
 after\_covid19 = ifelse(YEAR > 2020 | (YEAR == 2020 & MONTH >= 3), 1, 0))  
  
# Merge YEAR & MONTH variable, generate the new variable named date  
  
data9 <- data8 %>%  
 mutate(date = paste(YEAR, MONTH, sep = "-"))

## Regression model

#### Q1: How has COVID affected the health of the retail industry, as measured by employment?

# Use data9 for model1  
# DID Regression model with fixed effect  
# regression model 1: Y is unemployent rate, X is retail industry, Z is time after Covid-19  
  
model1 <- feols(unemployment ~ after\_covid19 + retail + after\_covid19\*retail, data = data9, vcov = 'hetero')  
summary(model1)

OLS estimation, Dep. Var.: unemployment  
Observations: 1,122,389   
Standard-errors: Heteroskedasticity-robust   
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 0.033468 0.000244 137.12112 < 2.2e-16 \*\*\*  
after\_covid19 0.041745 0.000457 91.27182 < 2.2e-16 \*\*\*  
retail 0.010245 0.000862 11.88817 < 2.2e-16 \*\*\*  
after\_covid19:retail -0.000832 0.001532 -0.54322 0.58698   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
RMSE: 0.224537 Adj. R2: 0.00866

# Display result of regression model 1  
etable(model1)

model1  
Dependent Var.: unemployment  
   
Constant 0.0335\*\*\* (0.0002)  
after\_covid19 0.0417\*\*\* (0.0005)  
retail 0.0102\*\*\* (0.0009)  
after\_covid19 x retail -0.0008 (0.0015)  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
S.E. type Heteroskedas.-rob.  
Observations 1,122,389  
R2 0.00866  
Adj. R2 0.00866  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Model 1:

after\_covid19 0.0417\*\*\* (0.0005)

retail 0.0102\*\*\* (0.0009)

after\_covid19 x retail -0.0008 (0.0015)

**Interpretation of Results**

First, it’s important to note that because our unemployment variable is categorical, we’re not interpreting its predicted values as the unemployment rate, but rather the probability of being unemployed. The unemployment rate is a rigorous economic indicator, which was not collected by the CPS.

After the onset of the pandemic, the average probability of being unemployed across all industries increased 4.2% (0.0417) compared to before the pandemic. Individuals in the retail industry experienced an average probability of being unemployed 1% higher (0.0102) than those in other industries. The probability of individuals in the retail industry being unemployed before the pandemic was 4.3% (0.0429) and 8.5% (0.0846) after the pandemic.

# Use data9 for model2  
# DID Regression model with fixed effects  
#log(FAMINC) = log(Family income of householder)  
model2 <- feols(log(FAMINC) ~ after\_covid19 + retail + after\_covid19\*retail, data = data9, vcov = 'hetero')  
summary(model2)

OLS estimation, Dep. Var.: log(FAMINC)  
Observations: 1,122,389   
Standard-errors: Heteroskedasticity-robust   
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 6.630047 0.000394 16848.81158 < 2.2e-16 \*\*\*  
after\_covid19 0.014266 0.000551 25.86837 < 2.2e-16 \*\*\*  
retail -0.036435 0.001387 -26.26882 < 2.2e-16 \*\*\*  
after\_covid19:retail 0.005851 0.001903 3.07495 0.0021054 \*\*   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
RMSE: 0.281505 Adj. R2: 0.002005

# Display result of regression model 2  
  
etable(model2)

model2  
Dependent Var.: log(FAMINC)  
   
Constant 6.630\*\*\* (0.0004)  
after\_covid19 0.0143\*\*\* (0.0006)  
retail -0.0364\*\*\* (0.0014)  
after\_covid19 x retail 0.0059\*\* (0.0019)  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
S.E. type Heteroskedast.-rob.  
Observations 1,122,389  
R2 0.00201  
Adj. R2 0.00200  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Model 2：

after\_covid19 0.0143\*\*\* (0.0006)

retail -0.0364\*\*\* (0.0014)

after\_covid19 x retail 0.0059\*\* (0.0019)

In addition to the unemployment rate, we wanted to analyze the impact of the pandemic on family income because of potential shifts between full-time and part-time employment as well as hiring or promotion freezes and government stimulus checks. After the onset of the pandemic, the average household income across industries increased by 1.4% (0.0143) compared to before the pandemic. Individuals in the retail industry experienced an average household income 3.6% lower (-0.0364) than those in other industries. Comparing the average household income across other industries, we found that the pandemic had an additional impact on the household income of those in the retail industry.

#### Q2: How has retail fared relative to other industries?

# data10 is only used in Regression Model 3 & 4 !  
# Removing rows where the "IND" (Industry code) variable is equal to 0   
  
data10 <- subset(data9, IND != 0)  
  
model3 <- feols(unemployment ~ after\_covid19 + wholesale + ms + manufacturing + service + agriculture + military + hospitality + retail + wholesale \* after\_covid19 + ms \* after\_covid19 + manufacturing \* after\_covid19 + service \* after\_covid19 + agriculture \* after\_covid19 + military \* after\_covid19 + hospitality\*after\_covid19 + retail\*after\_covid19, data = data10, vcov = 'hetero')

The variables 'retail' and 'after\_covid19:retail' have been removed because of collinearity (see $collin.var).

summary(model3)

OLS estimation, Dep. Var.: unemployment  
Observations: 1,110,438   
Standard-errors: Heteroskedasticity-robust   
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 0.043713 0.000827 52.888905 < 2.2e-16 \*\*\*  
after\_covid19 0.040912 0.001462 27.987018 < 2.2e-16 \*\*\*  
wholesale -0.017708 0.001617 -10.951145 < 2.2e-16 \*\*\*  
ms -0.015203 0.000977 -15.555025 < 2.2e-16 \*\*\*  
manufacturing -0.012955 0.001088 -11.904882 < 2.2e-16 \*\*\*  
service -0.016754 0.000879 -19.062682 < 2.2e-16 \*\*\*  
agriculture 0.000064 0.001916 0.033194 9.7352e-01   
military 0.956287 0.000827 1157.031296 < 2.2e-16 \*\*\*  
hospitality 0.009043 0.001266 7.145101 8.9985e-13 \*\*\*  
after\_covid19:wholesale -0.013928 0.002916 -4.777043 1.7791e-06 \*\*\*  
after\_covid19:ms -0.017408 0.001720 -10.123716 < 2.2e-16 \*\*\*  
after\_covid19:manufacturing -0.008672 0.001957 -4.432347 9.3222e-06 \*\*\*  
after\_covid19:service -0.004430 0.001568 -2.824716 4.7323e-03 \*\*   
after\_covid19:agriculture -0.011455 0.003287 -3.484461 4.9315e-04 \*\*\*  
after\_covid19:military -0.040912 0.001462 -27.987018 < 2.2e-16 \*\*\*  
after\_covid19:hospitality 0.103524 0.002612 39.631120 < 2.2e-16 \*\*\*  
... 2 variables were removed because of collinearity (retail and after\_covid19:retail)  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
RMSE: 0.218745 Adj. R2: 0.025489

# Display result of regression model 3  
etable(model3)

model3  
Dependent Var.: unemployment  
   
Constant 0.0437\*\*\* (0.0008)  
after\_covid19 0.0409\*\*\* (0.0015)  
wholesale -0.0177\*\*\* (0.0016)  
ms -0.0152\*\*\* (0.0010)  
manufacturing -0.0129\*\*\* (0.0011)  
service -0.0168\*\*\* (0.0009)  
agriculture 6.36e-5 (0.0019)  
military 0.9563\*\*\* (0.0008)  
hospitality 0.0090\*\*\* (0.0013)  
after\_covid19 x wholesale -0.0139\*\*\* (0.0029)  
after\_covid19 x ms -0.0174\*\*\* (0.0017)  
after\_covid19 x manufacturing -0.0087\*\*\* (0.0020)  
after\_covid19 x service -0.0044\*\* (0.0016)  
after\_covid19 x agriculture -0.0114\*\*\* (0.0033)  
after\_covid19 x military -0.0409\*\*\* (0.0015)  
after\_covid19 x hospitality 0.1035\*\*\* (0.0026)  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
S.E. type Heteroskedast.-rob.  
Observations 1,110,438  
R2 0.02550  
Adj. R2 0.02549  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Model 3：

after\_covid19 0.0409\*\*\* (0.0015)

wholesale -0.0177\*\*\* (0.0016)

ms -0.0152\*\*\* (0.0010)

manufacturing -0.0129\*\*\* (0.0011)

service -0.0168\*\*\* (0.0009)

agriculture 6.36e-5 (0.0019)

military 0.9563\*\*\* (0.0008)

hospitality 0.0090\*\*\* (0.0013)

after\_covid19 x wholesale -0.0139\*\*\* (0.0029)

after\_covid19 x ms -0.0174\*\*\* (0.0017)

after\_covid19 x manufacturing -0.0087\*\*\* (0.0020)

after\_covid19 x service -0.0044\*\* (0.0016)

after\_covid19 x agriculture -0.0114\*\*\* (0.0033)

after\_covid19 x military -0.0409\*\*\* (0.0015)

after\_covid19 x hospitality 0.1035\*\*\* (0.0026)

Comparing the impact of the pandemic on the retail industry relative to other industries, our regression coefficients indicated that after the pandemic, the probability of being unemployed in wholesale is 3.2 percentages points lower than in retail (-0.0177 + -0.0139), suggesting that wholesale experiences less impact from the pandemic compared to retail. The coefficients for management & science (-3.3) , manufacturing (-2.2), service (-2.1), and agriculture (-1.1) were all negative. This suggested that all of these industries are also less affected by the pandemic compared to retail. The hospitality and military coefficients are positive, which means they were more impacted by the pandemic compared to retail. The probability of unemployment is 11.3 percentage points higher in the hospitality industry and 95.6 percentage points higher in the military industry compared to retail after the pandemic.

# data10 is only used in Regression Model 3 & 4 !  
#log(FAMINC) = log(Family income of householder)  
  
model4 <- feols(log(FAMINC) ~ after\_covid19 + wholesale + ms + manufacturing + service + agriculture + military + hospitality + retail + wholesale \* after\_covid19 + ms \* after\_covid19 + manufacturing \* after\_covid19 + service \* after\_covid19 + agriculture \* after\_covid19 + military \* after\_covid19 + hospitality\*after\_covid19 + retail\*after\_covid19, data = data10, vcov = 'hetero')

The variables 'retail' and 'after\_covid19:retail' have been removed because of collinearity (see $collin.var).

summary(model4)

OLS estimation, Dep. Var.: log(FAMINC)  
Observations: 1,110,438   
Standard-errors: Heteroskedasticity-robust   
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 6.593612 0.001330 4957.451888 < 2.2e-16 \*\*\*  
after\_covid19 0.020117 0.001821 11.046936 < 2.2e-16 \*\*\*  
wholesale 0.059585 0.002529 23.557404 < 2.2e-16 \*\*\*  
ms 0.059880 0.001560 38.375901 < 2.2e-16 \*\*\*  
manufacturing 0.045103 0.001721 26.212584 < 2.2e-16 \*\*\*  
service 0.043534 0.001425 30.559768 < 2.2e-16 \*\*\*  
agriculture 0.004482 0.003110 1.441073 0.14956442   
military -0.112381 0.052605 -2.136301 0.03265510 \*   
hospitality -0.044377 0.002125 -20.887445 < 2.2e-16 \*\*\*  
after\_covid19:wholesale -0.006935 0.003433 -2.020218 0.04336104 \*   
after\_covid19:ms -0.009160 0.002147 -4.266140 0.00001989 \*\*\*  
after\_covid19:manufacturing -0.007204 0.002384 -3.021494 0.00251536 \*\*   
after\_covid19:service -0.006802 0.001956 -3.477824 0.00050552 \*\*\*  
after\_covid19:agriculture 0.002218 0.004286 0.517556 0.60476806   
after\_covid19:military 0.064073 0.070121 0.913751 0.36084787   
after\_covid19:hospitality -0.007350 0.003042 -2.415861 0.01569823 \*   
... 2 variables were removed because of collinearity (retail and after\_covid19:retail)  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
RMSE: 0.278774 Adj. R2: 0.011217

# Display result of regression model 4  
  
etable(model4)

model4  
Dependent Var.: log(FAMINC)  
   
Constant 6.594\*\*\* (0.0013)  
after\_covid19 0.0201\*\*\* (0.0018)  
wholesale 0.0596\*\*\* (0.0025)  
ms 0.0599\*\*\* (0.0016)  
manufacturing 0.0451\*\*\* (0.0017)  
service 0.0435\*\*\* (0.0014)  
agriculture 0.0045 (0.0031)  
military -0.1124\* (0.0526)  
hospitality -0.0444\*\*\* (0.0021)  
after\_covid19 x wholesale -0.0069\* (0.0034)  
after\_covid19 x ms -0.0092\*\*\* (0.0021)  
after\_covid19 x manufacturing -0.0072\*\* (0.0024)  
after\_covid19 x service -0.0068\*\*\* (0.0020)  
after\_covid19 x agriculture 0.0022 (0.0043)  
after\_covid19 x military 0.0641 (0.0701)  
after\_covid19 x hospitality -0.0073\* (0.0030)  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
S.E. type Heteroskedast.-rob.  
Observations 1,110,438  
R2 0.01123  
Adj. R2 0.01122  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Model 4：(We suggest not retaining model 4)

after\_covid19 x ms -0.0092\*\*\* (0.0021)

after\_covid19 x manufacturing -0.0072\*\* (0.0024)

after\_covid19 x service -0.0068\*\*\* (0.0020)

After the pandemic, the average household incomes in the management & science , manufacturing, and service industries are lower than in the retail industry. The coefficients are negative and significant.

after\_covid19 x agriculture 0.0022 (0.0043)

after\_covid19 x military 0.0641 (0.0701)

However, other industries are not significant, so there is no need to discuss them.

# data11 is only used in Model 5 !!  
# make date factor variable   
# date is stored as a character variable  
  
data11 <- data10  
data11$date <- factor(data11$date)  
  
  
#setting the reference group is 2020-2, run regression analysis with the specified reference group.  
  
data11$date <- relevel(data11$date, ref = "2020-2")  
model5 <- feols(unemployment ~ retail \* date, data = data11, vcov = 'hetero')  
summary(model5)

OLS estimation, Dep. Var.: unemployment  
Observations: 1,110,438   
Standard-errors: Heteroskedasticity-robust   
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 0.034061 0.000857 39.732932 < 2.2e-16 \*\*\*  
retail 0.009943 0.003012 3.300931 9.6367e-04 \*\*\*  
date2019-10 -0.006148 0.001158 -5.309818 1.0976e-07 \*\*\*  
date2019-11 -0.005572 0.001163 -4.790000 1.6680e-06 \*\*\*  
date2019-12 -0.004433 0.001175 -3.771790 1.6209e-04 \*\*\*  
date2019-3 -0.000689 0.001206 -0.570823 5.6812e-01   
date2019-4 -0.005733 0.001158 -4.949485 7.4421e-07 \*\*\*  
date2019-5 -0.005785 0.001163 -4.973422 6.5791e-07 \*\*\*  
date2019-6 -0.002152 0.001193 -1.803665 7.1284e-02 .   
date2019-7 -0.001144 0.001205 -0.949581 3.4233e-01   
date2019-8 -0.001715 0.001196 -1.434192 1.5152e-01   
date2019-9 -0.005464 0.001164 -4.695178 2.6641e-06 \*\*\*  
date2020-1 0.000407 0.001220 0.333637 7.3865e-01   
date2020-10 0.024599 0.001432 17.178247 < 2.2e-16 \*\*\*  
date2020-11 0.022184 0.001426 15.556513 < 2.2e-16 \*\*\*  
date2020-12 0.023255 0.001450 16.034658 < 2.2e-16 \*\*\*  
date2020-3 0.005961 0.001311 4.546010 5.4678e-06 \*\*\*  
date2020-4 0.094818 0.001955 48.491666 < 2.2e-16 \*\*\*  
date2020-5 0.076748 0.001879 40.839486 < 2.2e-16 \*\*\*  
date2020-6 0.060933 0.001809 33.681759 < 2.2e-16 \*\*\*  
date2020-7 0.055579 0.001755 31.671576 < 2.2e-16 \*\*\*  
date2020-8 0.038984 0.001609 24.223258 < 2.2e-16 \*\*\*  
date2020-9 0.034108 0.001515 22.510228 < 2.2e-16 \*\*\*  
date2021-1 0.026233 0.001461 17.956098 < 2.2e-16 \*\*\*  
date2021-2 0.025036 0.001457 17.188808 < 2.2e-16 \*\*\*  
retail:date2019-10 0.002589 0.004137 0.625756 5.3147e-01   
retail:date2019-11 -0.003068 0.004049 -0.757777 4.4858e-01   
retail:date2019-12 -0.000469 0.004122 -0.113869 9.0934e-01   
retail:date2019-3 0.007196 0.004384 1.641300 1.0074e-01   
retail:date2019-4 0.008402 0.004303 1.952557 5.0872e-02 .   
retail:date2019-5 0.007653 0.004270 1.792032 7.3128e-02 .   
retail:date2019-6 -0.001426 0.004173 -0.341622 7.3264e-01   
retail:date2019-7 0.003858 0.004309 0.895355 3.7060e-01   
retail:date2019-8 0.004997 0.004301 1.161912 2.4527e-01   
retail:date2019-9 0.001659 0.004168 0.398147 6.9052e-01   
retail:date2020-1 0.003609 0.004335 0.832652 4.0504e-01   
retail:date2020-10 -0.000276 0.004803 -0.057486 9.5416e-01   
retail:date2020-11 -0.008853 0.004589 -1.929200 5.3706e-02 .   
retail:date2020-12 -0.011369 0.004627 -2.457227 1.4002e-02 \*   
retail:date2020-3 -0.002062 0.004529 -0.455278 6.4891e-01   
retail:date2020-4 0.032321 0.006841 4.724719 2.3046e-06 \*\*\*  
retail:date2020-5 0.027593 0.006618 4.169317 3.0554e-05 \*\*\*  
retail:date2020-6 0.007128 0.006188 1.151860 2.4938e-01   
retail:date2020-7 -0.006643 0.005750 -1.155351 2.4795e-01   
retail:date2020-8 -0.002315 0.005411 -0.427851 6.6876e-01   
retail:date2020-9 -0.003240 0.005041 -0.642786 5.2036e-01   
retail:date2021-1 -0.001116 0.004869 -0.229109 8.1878e-01   
retail:date2021-2 -0.002443 0.004850 -0.503732 6.1445e-01   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
RMSE: 0.219829 Adj. R2: 0.01578

### Model 5：

To determine pandemic recovery and the impact of the pandemic over time, using February 2020 as a reference, the probability of being unemployed was generally lower before February 2020. In March 2020, the probability of being unemployed increased by 0.6% (0.006) compared to February 2020. In April 2020, the probability of being unemployed increased by 9.5% (0.0948) compared to February 2020. Until February 2021, the coefficient for the date variable kept decreasing, but it remained positive, indicating a decrease in the probability of being unemployed and an improvement in the economy. However, it still remained higher than in February 2020, indicating that the economy had not yet recovered to pre-pandemic levels.

# Display result of regression model 5  
etable(model5)

model5  
Dependent Var.: unemployment  
   
Constant 0.0341\*\*\* (0.0009)  
retail 0.0099\*\*\* (0.0030)  
date2019-10 -0.0061\*\*\* (0.0012)  
date2019-11 -0.0056\*\*\* (0.0012)  
date2019-12 -0.0044\*\*\* (0.0012)  
date2019-3 -0.0007 (0.0012)  
date2019-4 -0.0057\*\*\* (0.0012)  
date2019-5 -0.0058\*\*\* (0.0012)  
date2019-6 -0.0022. (0.0012)  
date2019-7 -0.0011 (0.0012)  
date2019-8 -0.0017 (0.0012)  
date2019-9 -0.0055\*\*\* (0.0012)  
date2020-1 0.0004 (0.0012)  
date2020-10 0.0246\*\*\* (0.0014)  
date2020-11 0.0222\*\*\* (0.0014)  
date2020-12 0.0233\*\*\* (0.0014)  
date2020-3 0.0060\*\*\* (0.0013)  
date2020-4 0.0948\*\*\* (0.0020)  
date2020-5 0.0767\*\*\* (0.0019)  
date2020-6 0.0609\*\*\* (0.0018)  
date2020-7 0.0556\*\*\* (0.0018)  
date2020-8 0.0390\*\*\* (0.0016)  
date2020-9 0.0341\*\*\* (0.0015)  
date2021-1 0.0262\*\*\* (0.0015)  
date2021-2 0.0250\*\*\* (0.0015)  
retail x date2019-10 0.0026 (0.0041)  
retail x date2019-11 -0.0031 (0.0040)  
retail x date2019-12 -0.0005 (0.0041)  
retail x date2019-3 0.0072 (0.0044)  
retail x date2019-4 0.0084. (0.0043)  
retail x date2019-5 0.0077. (0.0043)  
retail x date2019-6 -0.0014 (0.0042)  
retail x date2019-7 0.0039 (0.0043)  
retail x date2019-8 0.0050 (0.0043)  
retail x date2019-9 0.0017 (0.0042)  
retail x date2020-1 0.0036 (0.0043)  
retail x date2020-10 -0.0003 (0.0048)  
retail x date2020-11 -0.0089. (0.0046)  
retail x date2020-12 -0.0114\* (0.0046)  
retail x date2020-3 -0.0021 (0.0045)  
retail x date2020-4 0.0323\*\*\* (0.0068)  
retail x date2020-5 0.0276\*\*\* (0.0066)  
retail x date2020-6 0.0071 (0.0062)  
retail x date2020-7 -0.0066 (0.0057)  
retail x date2020-8 -0.0023 (0.0054)  
retail x date2020-9 -0.0032 (0.0050)  
retail x date2021-1 -0.0011 (0.0049)  
retail x date2021-2 -0.0024 (0.0048)  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
S.E. type Heteroskedast.-rob.  
Observations 1,110,438  
R2 0.01582  
Adj. R2 0.01578  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

retail x date2020-4 0.0323\*\*\* (0.0068)

retail x date2020-5 0.0276\*\*\* (0.0066)

retail x date2020-12 -0.0114\* (0.0046)

The probability of being unemployed in the retail industry in April 2020 was 3.2% (0.0323) higher than in other industries, indicating that the retail industry was more severely affected by the pandemic. In May 2020 it was 2.8%(0.0276) higher than in other industries, however, in December 2020 was 1.1% (-0.0114) lower than in other industries, indicating that the retail industry recovered faster than other industries in that month. The coefficients after April 2020 are insignificant, indicating that in other months, the probability of being unemployed in the retail industry has not significantly increased or decreased compared to other industries.

#### Q3: Retail needs to worry about who has money to spend - what has changed about who is working and earning money?

# Use data11 for Model 6  
model6 <- feols(unemployment ~ after\_covid19 + black + aapi + indigenous + multiracial + other + white + black \* after\_covid19 + aapi \* after\_covid19 + indigenous \* after\_covid19 + multiracial \* after\_covid19 + other \* after\_covid19 + white \* after\_covid19, data = data9, vcov = 'hetero')

The variables 'other', 'white' and two others have been removed because of collinearity (see $collin.var).

summary(model6)

OLS estimation, Dep. Var.: unemployment  
Observations: 1,122,389   
Standard-errors: Heteroskedasticity-robust   
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 0.030496 0.000247 123.69510 < 2.2e-16 \*\*\*  
after\_covid19 0.038531 0.000464 83.12154 < 2.2e-16 \*\*\*  
black 0.031523 0.001008 31.28514 < 2.2e-16 \*\*\*  
aapi -0.003754 0.000862 -4.35651 1.3217e-05 \*\*\*  
indigenous 0.052155 0.003225 16.17112 < 2.2e-16 \*\*\*  
multiracial 0.023135 0.002132 10.85148 < 2.2e-16 \*\*\*  
after\_covid19:black 0.012438 0.001766 7.04187 1.8979e-12 \*\*\*  
after\_covid19:aapi 0.022674 0.001803 12.57652 < 2.2e-16 \*\*\*  
after\_covid19:indigenous 0.008303 0.005403 1.53668 1.2437e-01   
after\_covid19:multiracial 0.022185 0.003899 5.69076 1.2651e-08 \*\*\*  
... 4 variables were removed because of collinearity (other, white and 2 others [full set in $collin.var])  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
RMSE: 0.224151 Adj. R2: 0.012059

# update with other race variables

# Display result of regression model 6  
etable(model6)

model6  
Dependent Var.: unemployment  
   
Constant 0.0305\*\*\* (0.0002)  
after\_covid19 0.0385\*\*\* (0.0005)  
black 0.0315\*\*\* (0.0010)  
aapi -0.0038\*\*\* (0.0009)  
indigenous 0.0522\*\*\* (0.0032)  
multiracial 0.0231\*\*\* (0.0021)  
after\_covid19 x black 0.0124\*\*\* (0.0018)  
after\_covid19 x aapi 0.0227\*\*\* (0.0018)  
after\_covid19 x indigenous 0.0083 (0.0054)  
after\_covid19 x multiracial 0.0222\*\*\* (0.0039)  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
S.E. type Heteroskedast.-rob.  
Observations 1,122,389  
R2 0.01207  
Adj. R2 0.01206  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Model 6:

after\_covid19 x black 0.0124\*\*\* (0.0018)

after\_covid19 x aapi 0.0227\*\*\* (0.0018)

after\_covid19 x multiracial 0.0222\*\*\* (0.0039)

To further determine how the retail market may be impacted by the pandemic in terms of spending, we evaluated the probability of unemployment based on race, gender, and education level. The coefficients from our associated regression models, after the pandemic, showed a 1.2 percentage point increase in the probability of being unemployed for black individuals, a 2.3 percentage point increase for asian americans and pacific islanders, and a 2.2 percentage point increase for multiracial individuals compared to white individuals. The impact of the pandemic on the probability of being unemployed for indigenous individuals was not statistically significant.

# Use data10 for Model 7  
  
model7 <- feols(unemployment ~ after\_covid19 + female + female \* after\_covid19, data = data10, vcov = 'hetero')  
summary(model7)

OLS estimation, Dep. Var.: unemployment  
Observations: 1,110,438   
Standard-errors: Heteroskedasticity-robust   
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 0.032913 0.000320 102.82484 < 2.2e-16 \*\*\*  
after\_covid19 0.040770 0.000598 68.14739 < 2.2e-16 \*\*\*  
female -0.001549 0.000456 -3.39386 6.8918e-04 \*\*\*  
after\_covid19:female 0.003572 0.000865 4.12788 3.6615e-05 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
RMSE: 0.220573 Adj. R2: 0.009149

# Display result of regression model 8  
etable(model7)

model7  
Dependent Var.: unemployment  
   
Constant 0.0329\*\*\* (0.0003)  
after\_covid19 0.0408\*\*\* (0.0006)  
female -0.0015\*\*\* (0.0005)  
after\_covid19 x female 0.0036\*\*\* (0.0009)  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
S.E. type Heteroskedast.-rob.  
Observations 1,110,438  
R2 0.00915  
Adj. R2 0.00915  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Model 7:

female -0.0015\*\*\* (0.0005)

after\_covid19 x female 0.0036\*\*\* (0.0009)

The probability of unemployment for women decreased by 0.15 percentage points compared to men. This coefficient tells us that there is no evidence to suggest that after the pandemic, the change in the probability of being unemployed for women is higher than for men because the change is so small.

# Use data10 for Model 8  
model8 <- feols(unemployment ~ after\_covid19 + under\_hs + hs\_diploma + some\_college + associate + bachelor + graduate + under\_hs \* after\_covid19 + hs\_diploma \* after\_covid19 + some\_college \* after\_covid19 + associate \* after\_covid19 + bachelor \* after\_covid19 + graduate \* after\_covid19, data = data10, vcov = 'hetero')

The variables 'graduate' and 'after\_covid19:graduate' have been removed because of collinearity (see $collin.var).

summary(model8)

OLS estimation, Dep. Var.: unemployment  
Observations: 1,110,438   
Standard-errors: Heteroskedasticity-robust   
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 0.015237 0.000429 35.4991 < 2.2e-16 \*\*\*  
after\_covid19 0.017477 0.000777 22.4844 < 2.2e-16 \*\*\*  
under\_hs 0.047731 0.001289 37.0431 < 2.2e-16 \*\*\*  
hs\_diploma 0.028108 0.000666 42.1762 < 2.2e-16 \*\*\*  
some\_college 0.020792 0.000722 28.8168 < 2.2e-16 \*\*\*  
associate 0.010544 0.000749 14.0708 < 2.2e-16 \*\*\*  
bachelor 0.005991 0.000571 10.4993 < 2.2e-16 \*\*\*  
after\_covid19:under\_hs 0.046892 0.002390 19.6231 < 2.2e-16 \*\*\*  
after\_covid19:hs\_diploma 0.036658 0.001234 29.6991 < 2.2e-16 \*\*\*  
after\_covid19:some\_college 0.038121 0.001387 27.4889 < 2.2e-16 \*\*\*  
after\_covid19:associate 0.026390 0.001453 18.1591 < 2.2e-16 \*\*\*  
after\_covid19:bachelor 0.015772 0.001067 14.7855 < 2.2e-16 \*\*\*  
... 2 variables were removed because of collinearity (graduate and after\_covid19:graduate)  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
RMSE: 0.219614 Adj. R2: 0.017733

# Display result of regression model 8  
etable(model8)

model8  
Dependent Var.: unemployment  
   
Constant 0.0152\*\*\* (0.0004)  
after\_covid19 0.0175\*\*\* (0.0008)  
under\_hs 0.0477\*\*\* (0.0013)  
hs\_diploma 0.0281\*\*\* (0.0007)  
some\_college 0.0208\*\*\* (0.0007)  
associate 0.0105\*\*\* (0.0007)  
bachelor 0.0060\*\*\* (0.0006)  
after\_covid19 x under\_hs 0.0469\*\*\* (0.0024)  
after\_covid19 x hs\_diploma 0.0367\*\*\* (0.0012)  
after\_covid19 x some\_college 0.0381\*\*\* (0.0014)  
after\_covid19 x associate 0.0264\*\*\* (0.0015)  
after\_covid19 x bachelor 0.0158\*\*\* (0.0011)  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
S.E. type Heteroskedas.-rob.  
Observations 1,110,438  
R2 0.01774  
Adj. R2 0.01773  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Model 8:

after\_covid19 x under\_hs 0.0469\*\*\* (0.0024)

after\_covid19 x hs\_diploma 0.0367\*\*\* (0.0012)

after\_covid19 x some\_college 0.0381\*\*\* (0.0014)

after\_covid19 x associate 0.0264\*\*\* (0.0015)

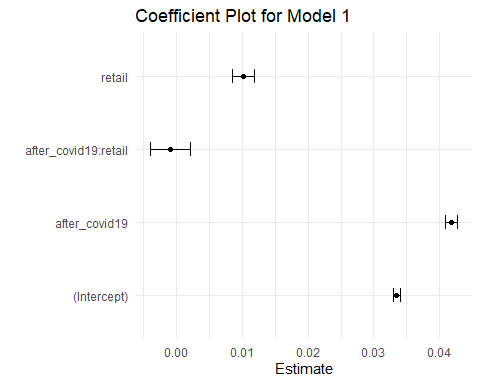
after\_covid19 x bachelor 0.0158\*\*\* (0.0011)

The probability of being unemployed for individuals with less than a high school diploma, after the pandemic, was 4.7 percentage points (0.0469) higher than those with graduate degrees. From there, the coefficient decreased for every additional level of education obtained by an individual (those with a high school diploma, some college, an associate degree, and a bachelor degree), suggesting that the impact of the pandemic on the probability of being unemployed decreases with more education.

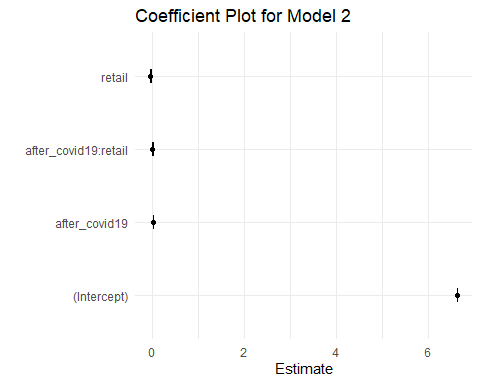
### Graphical Analysis

### Coefficient Plots

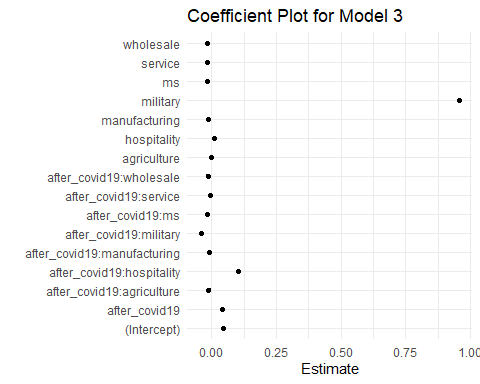
# Function to create coefficient plots  
create\_coefficient\_plot <- function(model, model\_name) {  
 # Get coefficients and confidence intervals directly from the model  
 ests <- coef(model)  
 confints <- confint(model)  
  
 # Create a data frame for plotting; use names() to get variable names  
 coef\_df <- data.frame(  
 term = names(ests),  
 estimate = ests,  
 conf.low = confints[, 1],  
 conf.high = confints[, 2]  
 )  
   
 # Plot using ggplot2  
 p <- ggplot(coef\_df, aes(x = estimate, y = term)) +  
 geom\_point() +  
 geom\_errorbarh(aes(xmin = conf.low, xmax = conf.high), height = 0.2) +  
 labs(title = paste("Coefficient Plot for", model\_name),  
 x = "Estimate", y = "") +  
 theme\_minimal()  
   
 print(p)  
}  
  
create\_coefficient\_plot(model1, "Model 1")



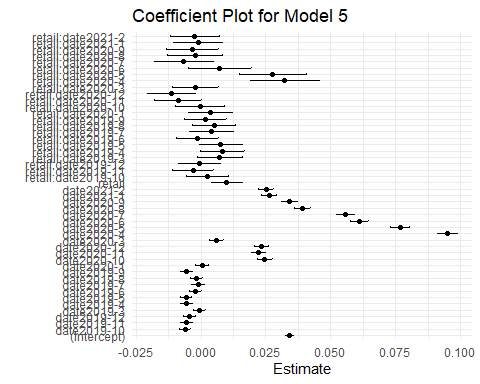
create\_coefficient\_plot(model2, "Model 2")



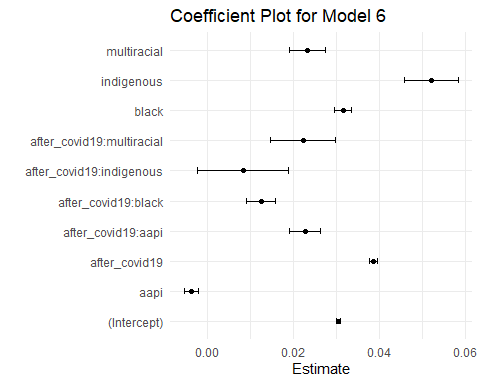
create\_coefficient\_plot(model3, "Model 3")



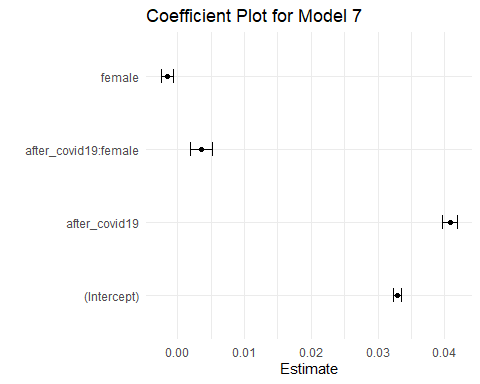
create\_coefficient\_plot(model5, "Model 5")



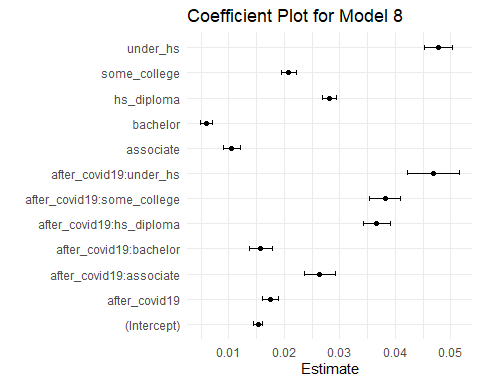
create\_coefficient\_plot(model6, "Model 6")



create\_coefficient\_plot(model7, "Model 7")



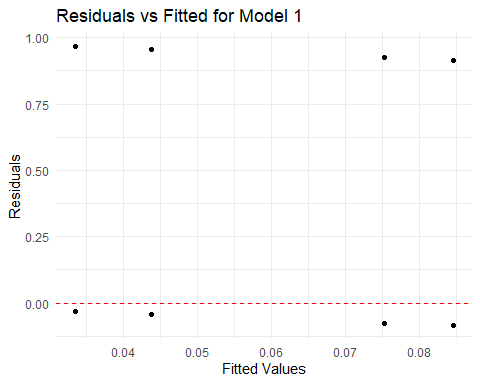
create\_coefficient\_plot(model8, "Model 8")



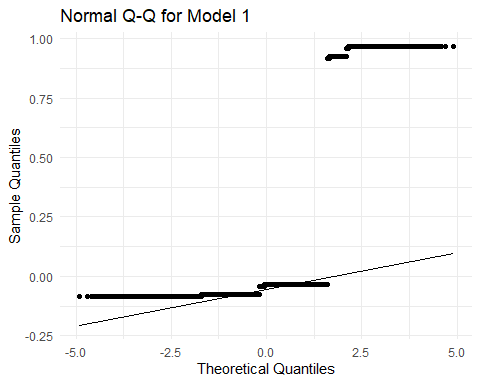
Coefficient plots visualize the estimated effects (coefficients) of each predictor variable in each regression model, including confidence intervals. Each point on the plot represents the estimated effect of one predictor on the dependent variable, while the horizontal lines represent the range of the confidence interval for that estimate. These plots show at a glance which variables have a statistically significant effect on the outcome, as indicated by confidence intervals that do not cross the vertical line at zero. They also allow for easy comparison of the magnitude and direction (positive or negative) of effects across variables. These coefficient plots condense complex information from models with multiple predictors into an easy to interpret visual format. By examining these plots, one can identify which variables play a significant role in the model and whether their effect is positive or negative.

### Diagnostic Check Plots

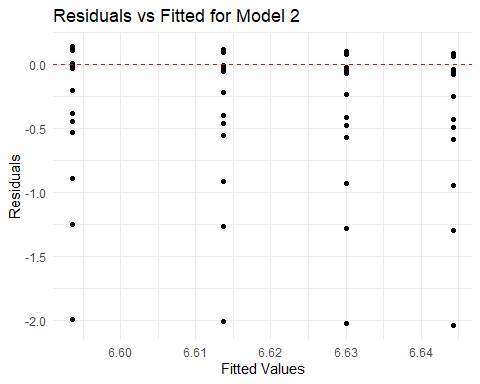
# Create a data frame with residuals and fitted values for Model 1  
model1\_df <- data.frame(  
 fitted\_values = fitted(model1),  
 residuals = resid(model1)  
)  
  
# Residuals vs Fitted Plot for Model 1  
ggplot(model1\_df, aes(x = fitted\_values, y = residuals)) +  
 geom\_point() +  
 geom\_hline(yintercept = 0, linetype = "dashed", color = "red") +  
 labs(title = "Residuals vs Fitted for Model 1", x = "Fitted Values", y = "Residuals") +  
 theme\_minimal()



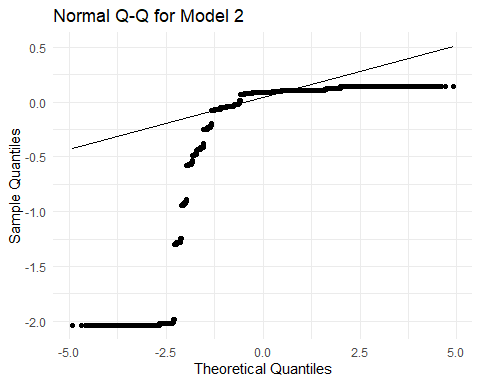
# For the Normal Q-Q plot, we'll use the "resid" function directly within the ggplot call  
ggplot(model1\_df, aes(sample = residuals)) +  
 stat\_qq() +  
 stat\_qq\_line() +  
 labs(title = "Normal Q-Q for Model 1", x = "Theoretical Quantiles", y = "Sample Quantiles") +  
 theme\_minimal()



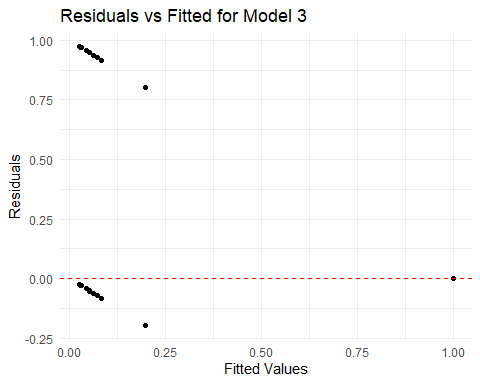
# Creating a data frame for Model 2  
model2\_df <- data.frame(  
 fitted\_values = fitted(model2),  
 residuals = resid(model2)  
)  
  
# Residuals vs Fitted Plot for Model 2  
ggplot(model2\_df, aes(x = fitted\_values, y = residuals)) +  
 geom\_point() +  
 geom\_hline(yintercept = 0, linetype = "dashed", color = "red") +  
 labs(title = "Residuals vs Fitted for Model 2", x = "Fitted Values", y = "Residuals") +  
 theme\_minimal()



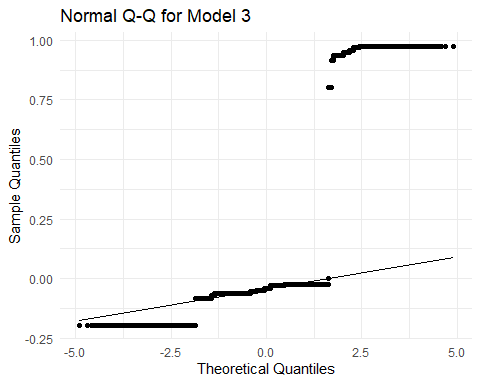
# Normal Q-Q Plot for Model 2  
ggplot(model2\_df, aes(sample = residuals)) +  
 stat\_qq() +  
 stat\_qq\_line() +  
 labs(title = "Normal Q-Q for Model 2", x = "Theoretical Quantiles", y = "Sample Quantiles") +  
 theme\_minimal()



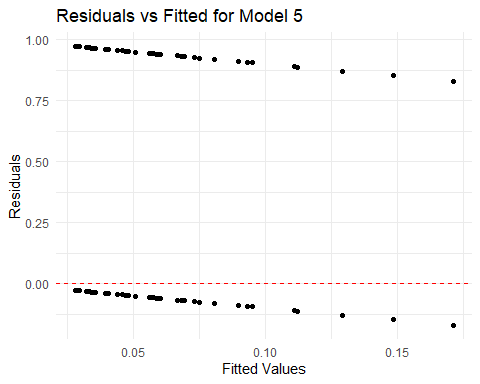
# Creating a data frame for Model 3  
model3\_df <- data.frame(  
 fitted\_values = fitted(model3),  
 residuals = resid(model3)  
)  
  
# Residuals vs Fitted Plot for Model 3  
ggplot(model3\_df, aes(x = fitted\_values, y = residuals)) +  
 geom\_point() +  
 geom\_hline(yintercept = 0, linetype = "dashed", color = "red") +  
 labs(title = "Residuals vs Fitted for Model 3", x = "Fitted Values", y = "Residuals") +  
 theme\_minimal()



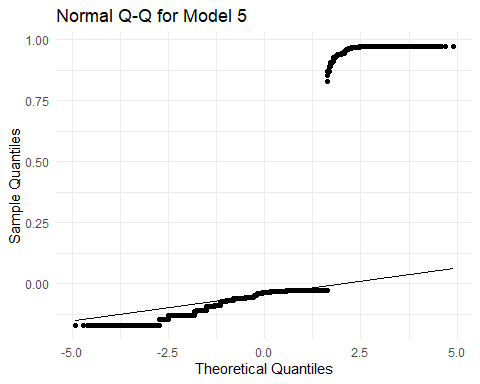
# Normal Q-Q Plot for Model 3  
ggplot(model3\_df, aes(sample = residuals)) +  
 stat\_qq() +  
 stat\_qq\_line() +  
 labs(title = "Normal Q-Q for Model 3", x = "Theoretical Quantiles", y = "Sample Quantiles") +  
 theme\_minimal()



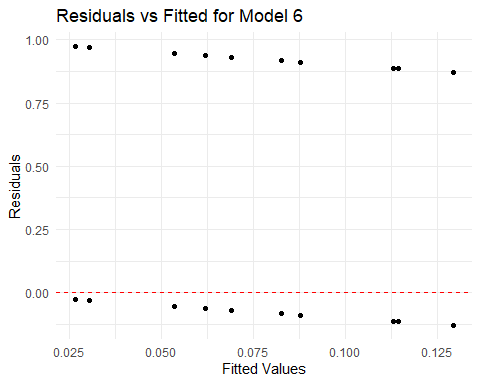
# Creating a data frame for Model 5  
model5\_df <- data.frame(  
 fitted\_values = fitted(model5),  
 residuals = resid(model5)  
)  
  
# Residuals vs Fitted Plot for Model 5  
ggplot(model5\_df, aes(x = fitted\_values, y = residuals)) +  
 geom\_point() +  
 geom\_hline(yintercept = 0, linetype = "dashed", color = "red") +  
 labs(title = "Residuals vs Fitted for Model 5", x = "Fitted Values", y = "Residuals") +  
 theme\_minimal()



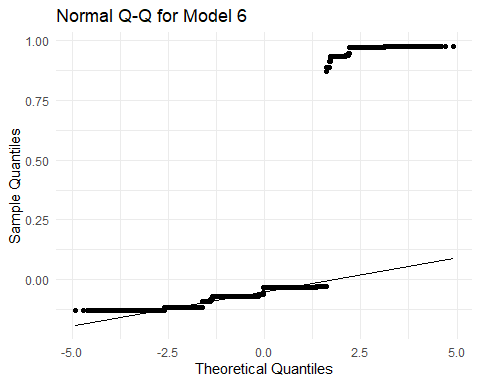
# Normal Q-Q Plot for Model 5  
ggplot(model5\_df, aes(sample = residuals)) +  
 stat\_qq() +  
 stat\_qq\_line() +  
 labs(title = "Normal Q-Q for Model 5", x = "Theoretical Quantiles", y = "Sample Quantiles") +  
 theme\_minimal()



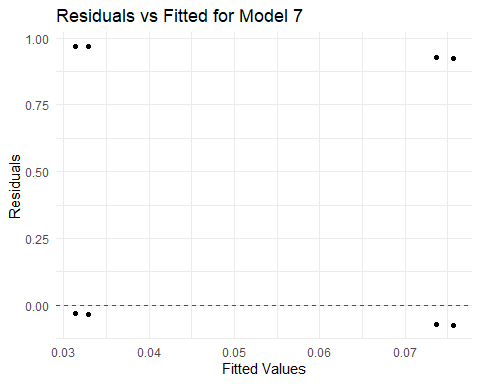
# Creating a data frame for Model 6  
model6\_df <- data.frame(  
 fitted\_values = fitted(model6),  
 residuals = resid(model6)  
)  
  
# Residuals vs Fitted Plot for Model 6  
ggplot(model6\_df, aes(x = fitted\_values, y = residuals)) +  
 geom\_point() +  
 geom\_hline(yintercept = 0, linetype = "dashed", color = "red") +  
 labs(title = "Residuals vs Fitted for Model 6", x = "Fitted Values", y = "Residuals") +  
 theme\_minimal()



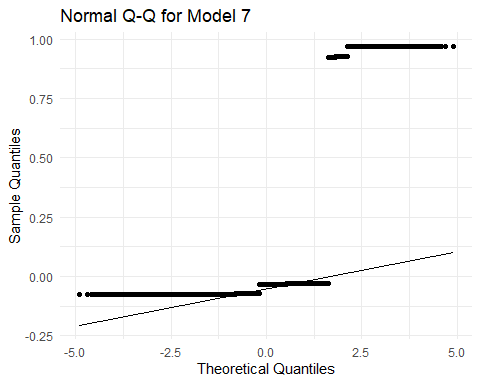
# Normal Q-Q Plot for Model 6  
ggplot(model6\_df, aes(sample = residuals)) +  
 stat\_qq() +  
 stat\_qq\_line() +  
 labs(title = "Normal Q-Q for Model 6", x = "Theoretical Quantiles", y = "Sample Quantiles") +  
 theme\_minimal()



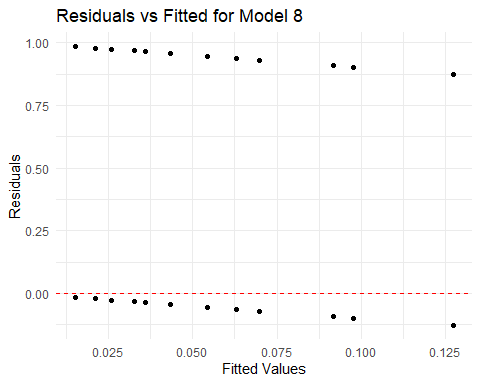
# Creating a data frame for Model 7  
model7\_df <- data.frame(  
 fitted\_values = fitted(model7),  
 residuals = resid(model7)  
)  
  
# Residuals vs Fitted Plot for Model 7  
ggplot(model7\_df, aes(x = fitted\_values, y = residuals)) +  
 geom\_point() +  
 geom\_hline(yintercept = 0, linetype = "dashed", color = "red") +  
 labs(title = "Residuals vs Fitted for Model 7", x = "Fitted Values", y = "Residuals") +  
 theme\_minimal()



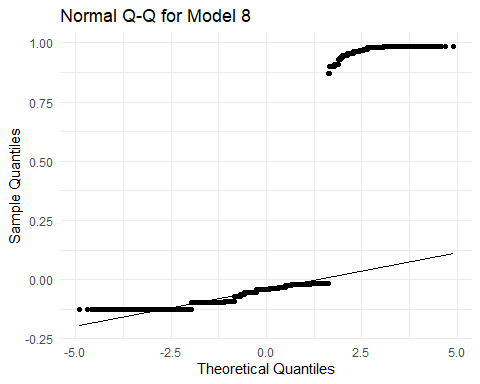
# Normal Q-Q Plot for Model 7  
ggplot(model7\_df, aes(sample = residuals)) +  
 stat\_qq() +  
 stat\_qq\_line() +  
 labs(title = "Normal Q-Q for Model 7", x = "Theoretical Quantiles", y = "Sample Quantiles") +  
 theme\_minimal()



# Creating a data frame for Model 8  
model8\_df <- data.frame(  
 fitted\_values = fitted(model8),  
 residuals = resid(model8)  
)  
  
# Residuals vs Fitted Plot for Model 8  
ggplot(model8\_df, aes(x = fitted\_values, y = residuals)) +  
 geom\_point() +  
 geom\_hline(yintercept = 0, linetype = "dashed", color = "red") +  
 labs(title = "Residuals vs Fitted for Model 8", x = "Fitted Values", y = "Residuals") +  
 theme\_minimal()



# Normal Q-Q Plot for Model 8  
ggplot(model8\_df, aes(sample = residuals)) +  
 stat\_qq() +  
 stat\_qq\_line() +  
 labs(title = "Normal Q-Q for Model 8", x = "Theoretical Quantiles", y = "Sample Quantiles") +  
 theme\_minimal()



We decided to employ two different types of plots to perform diagnostic checks, so that we can see if our regression models meet the assumptions for regression.

The first type are “residuals vs fitted” plots, which show the residuals (differences between observed and predicted values) on the y-axis against the predicted (or fitted) values on the x-axis. This type of plot helps check for the linearity and homoscedasticity assumptions. They are essential for identifying whether the linear model is appropriate for the data or not. Any obvious patterns like a funnel-shape or a curve suggest that the assumptions are not satisfied, indicating that the model may not be the best fit for the data or that transformations may be necessary.

The second type are “normal Q-Q” plots, which compare the quantiles of the residuals from the regression models to the quantiles of a normal distribution. It plots the theoretical quantiles of the normal distribution on the x-axis and the ordered residuals on the y-axis. This type of plot checks for normality of residuals, which is a key assumption for the validity of various statistical tests, including those for significance of coefficients. If the residuals are normally distributed, the points on the Q-Q plot will closely follow the diagonal line. Deviations from this line at the ends suggest heavy-tailed or skewed distributions, while a deviation throughout that forms a curve indicates that the residuals have a distribution that is not normal.

### **Conclusion**

Our analysis of the Current Population Survey (CPS) data from IPUMS has illuminated the impact of the COVID-19 pandemic on the retail industry, particularly in terms of employment, and offered comparisons with other sectors. This analysis also sheds light on the changing economic well-being of various demographic groups, which in turn affects their spending capabilities.

**COVID-19’s Impact on Retail Employment:** The pandemic led to a marked increase in unemployment within the retail sector, with the probability of unemployment rising by 1%. This reflects a significant disruption, as the overall likelihood of unemployment across industries increased by 4.2% following the pandemic’s onset.

**Retail’s Standing Relative to Other Industries:** The retail sector faced a steeper challenge than some industries but showed remarkable resilience, recovering faster by the end of 2020. While sectors like wholesale and manufacturing were less affected, the hospitality industry endured more severe impacts, highlighting the uneven distribution of the pandemic’s economic consequences across different fields.

**Shifts in Economic Well-being and Consumer Spending Power:** The pandemic disproportionately affected certain demographic groups, notably increasing unemployment probabilities among black, AAPI, and multiracial individuals compared to white individuals. Education emerged as a significant factor, with higher educational attainment correlating with lower unemployment risks. These changes indicate a shift in who has the financial means to spend, with implications for the retail sector’s focus on consumer engagement and market strategies.

In essence, while the COVID-19 pandemic has presented substantial challenges to the retail industry, it has also spurred adaptability and resilience. The differential impacts across demographic groups highlight the importance of nuanced market strategies for retailers aiming to engage a changing consumer base effectively.

### Summary

* The COVID-19 pandemic significantly disrupted the retail sector’s employment, illustrating the sector’s initial vulnerability and subsequent adaptability.
* Compared to other sectors, retail initially faced severe impacts but demonstrated a faster recovery, unlike the more heavily affected hospitality industry.
* The pandemic altered economic well-being across demographic groups, affecting consumer spending power. Those with higher education levels were less impacted, suggesting a potential shift in the retail market’s target demographics and spending patterns.