Data Translation Project

Task Outline

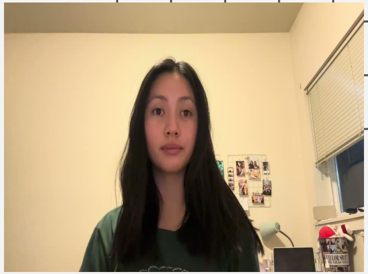
It’s April 2022, we work for a company in the retail sector. Our company knows how well they’re weathering the pandemic, but they are having difficulty figuring out what the effect has been on the rest of the retail sector and the rest of the economy. Everyone is being tight with their numbers!

So, they’ve pointed us towards the IPUMS release of the Current Population Survey, which the government often uses to understand changes in employment.

Why employment data? They figure that revenues can be misrepresented, or might be affected by government aid. But employment in an industry can tell you a lot about how well that industry is doing!

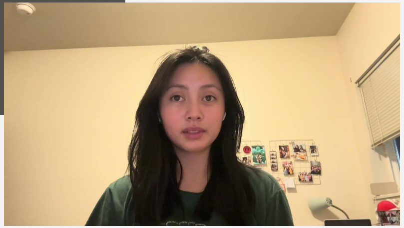
What questions will we be attempting to answer?

1. How has COVID affected the health of the retail industry, as measured by employment?
2. How has retail fared relative to other industries?

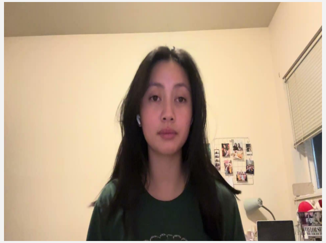


01

How has COVID affected the health of the retail industry?



What are we looking for?



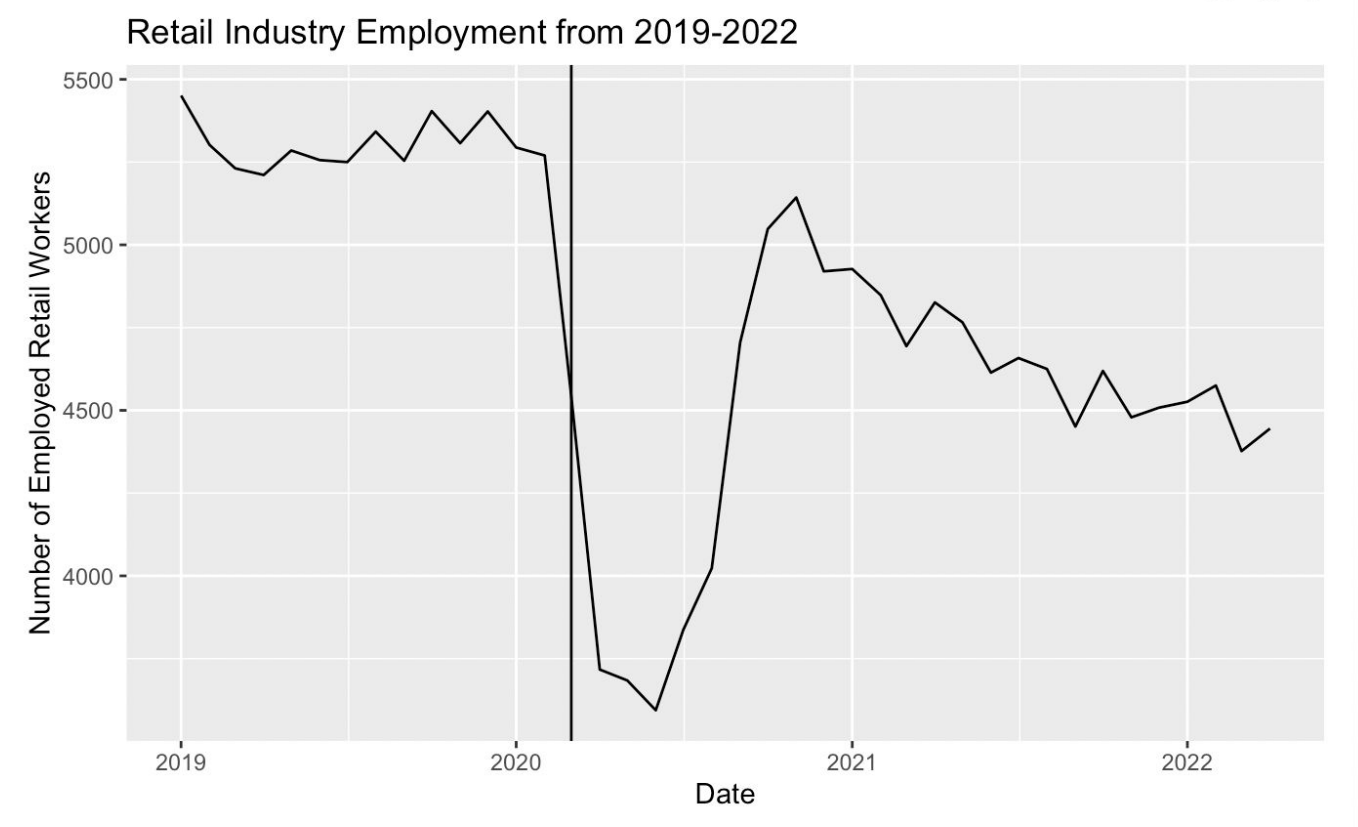
RELATIONSHIP

- **Outcome:** health of the retail industry
  - Measured by employment level
  - Looking at how employment of the retail industry changes
- **Treatment:** COVID
  - In the US, COVID lockdown was implemented in March 2020
  - Split data into two groups
    - pre-lockdown: March 2020 or prior
    - post-lockdown: After March 2020
- **Research Design:** <u>Regression Discontinuity Design</u>
  - Measures the impact of a treatment (COVID) on the basis of being above/below the cutoff value (March 2020) of a running variable (Date)

DATA

- **IPUMS release** of the **Current Population Survey**
  - Contains all types of data about the US population, including employment status, industry, sex, race, etc.
  - January 2019 - April 2022
  - Often used by government (credible)
  - Joined with ‘indnames.csv’ for industry names
- **Creating RDD**
  - Data filtered to only look at <u>retail</u> industry & dates from April 2019 to February 2022
  - *Variables:*
    - **Employment** = Total amount of employed people (aggregated to month level)
    - **COVID_center** = Date - 03/01/2020 (X - cutoff)
    - **post_lockdown** = TRUE if Date is past March 2020

Retail Employment Over Time



[redacted]

Regression Results

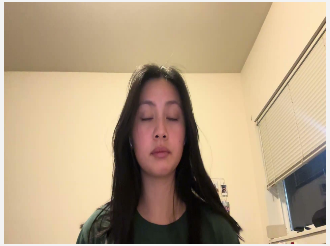
Interpretation: The retail industry after the COVID lockdown had 1703 less employed workers than before the COVID lockdown, and this result is statistically significant.

COVID had a negative impact on the health of the retail industry, dropping employment in the retail industry by 33.32%. While employment does seem to rise afterwards, employment levels haven't returned to the same level as pre-lockdown as of April 2022. This could be due to different policies implemented to mediate the impact of COVID, such as unemployment benefits or workers have moved onto other industries.

```
model1 <- feols(Employed ~ post_lockdown + COVID_center +
post_lockdown*COVID_center, data = joined_data_1, vcov = 'hetero')
etable(model1)
```

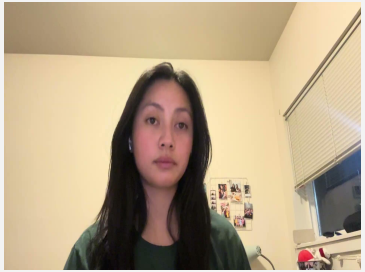
```
                                                model1
Dependent Var.:                               Employed

Constant                           5,112.8*** (198.5)
post_lockdownTRUE                 -1,703.8*** (249.4)
COVID_center                         -0.7300 (0.8529)
post_lockdownTRUE x COVID_center     6.150*** (1.200)
________________________________ ___________________
S.E. type                          Heteroskedast.-rob.
Observations                                        23
R2                                             0.83762
Adj. R2                                        0.81199
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



RDD Assumptions

- **Isolated Treatment:** the only factor changing at the cutoff is the treatment (COVID lockdown)
- **Window:** only interested data right above/below cutoff
  - April 2019 - February 2022
  - Small enough to focus on variation around COVID, closes other back doors
  - Big enough sample to properly estimate impact
- **Granular Running Variable:**
  - the employment data is aggregated at the month level (meaning that data is in chunks)
- **No Manipulation:** Companies were most likely unaware of COVID lockdown beforehand making manipulation implausible
- **Sharp Regression Discontinuity:** all data above the cutoff received the treatment (COVID lockdown) and all data below the cutoff did not receive the treatment



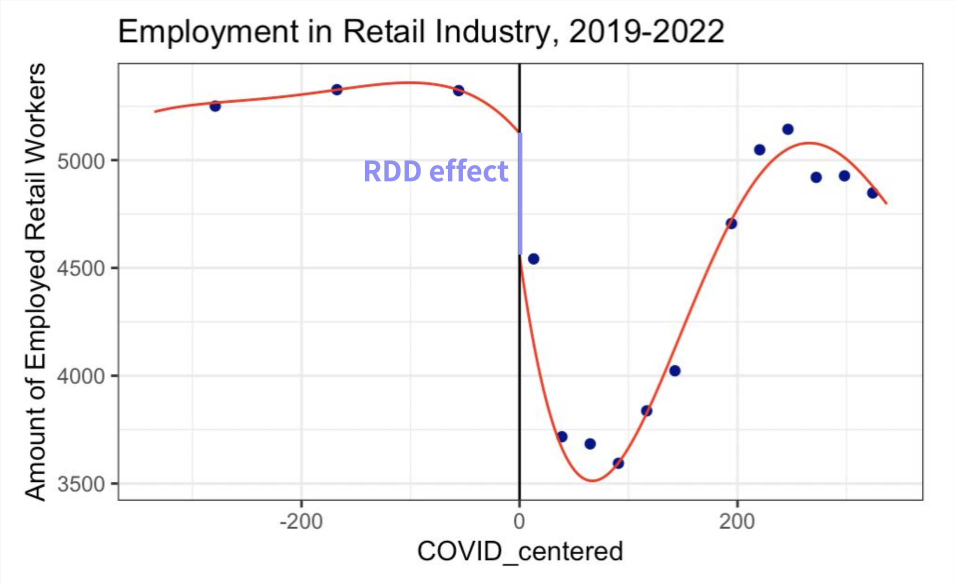
RDROBUST Data

Rdrobust package estimates RDD using local polynomial regression (less worry about linearity)

- Running the RDD: Retail industry experienced a 17.13% decrease in employment levels after COVID started.
  - Different results could be due to small sample
- rdplot allows us to better visualize the effect of COVID on employment

```
model1b <- rdrobust(joined_data_1$Employed, joined_data_1$COVID_center, c = 0,
kernel = 'uniform')
```

Method	Coef.	Std. Err.	z	P>|z|	[95% C.I.]
Conventional	-876.068	546.796	-1.602	0.109	[-1947.769 , 195.633]
Robust	-	-	-1.194	0.232	[-1884.272 , 457.507]

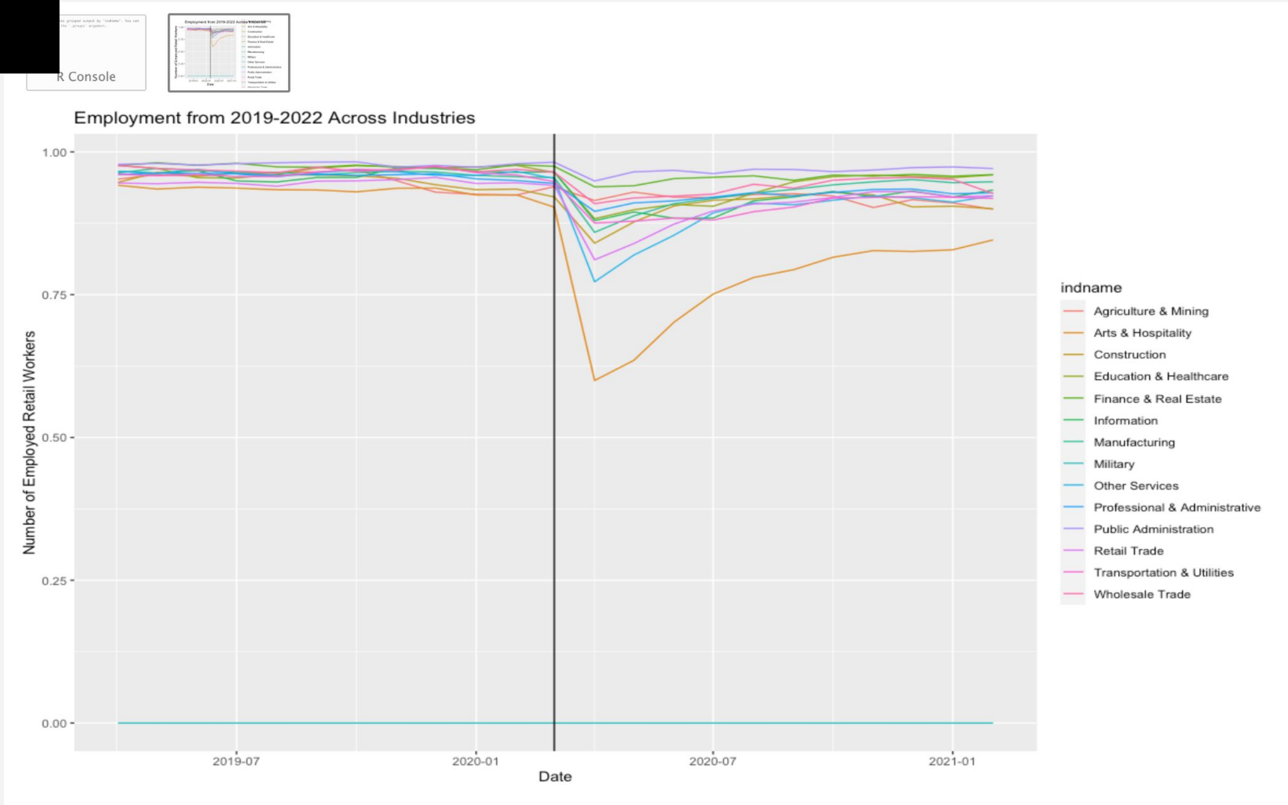


02

How has retail fared relative to other industries?

[redacted]

Graph Analysis



Regression Results

indname Coefficients: These show the difference in employment between each industry and the Retail Trade industry, which is the reference category. Negative coefficients for other industries suggest they have fared worse than retail in terms of employment numbers, and positive coefficients suggest they have fared better.

post_lockdownTRUE: This shows the change in employment in the retail industry after the lockdown was imposed. Since the coefficient is negative, it indicates a decrease in retail employment after the lockdown compared to before.

post_lockdownTRUE x COVID_center: The interaction term's positive coefficient suggests that while there was a drop in employment after the lockdown, the relative decline slowed over time.

```
                                                   model2
Dependent Var.:                                  Employed

Constant                              0.9412*** (0.0043)
post_lockdownTRUE                    -0.0795*** (0.0090)
COVID_center                         -2.43e-5. (1.43e-5)
indname = Agriculture&Mining            0.0114. (0.0059)
indname = Arts&Hospitality           -0.0716*** (0.0163)
indname = Construction                   0.0049 (0.0051)
indname = Education&Healthcare        0.0286*** (0.0043)
indname = Finance&RealEstate          0.0422*** (0.0050)
indname = Information                 0.0145*** (0.0041)
indname = Manufacturing               0.0218*** (0.0042)
indname = Military                   -0.9228*** (0.0068)
indname = OtherServices                  0.0033 (0.0068)
indname = Professional&Administrative 0.0181*** (0.0042)
indname = PublicAdministration        0.0501*** (0.0056)
indname = Transportation&Utilities      0.0102* (0.0041)
indname = WholesaleTrade              0.0303*** (0.0044)
post_lockdownTRUE x COVID_center     0.0002*** (3.75e-5)
_____________________________________ ___________________
S.E. type                            Heteroskedast.-rob.
Observations                                         322
R2                                               0.98884
Adj. R2                                          0.98826
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Overall Interpretation

The analyses show that retail employment experienced a significant decrease immediately following the lockdown (as indicated by the post_lockdownTRUE coefficient).

However, the positive and significant interaction term suggests that the rate of decline in retail employment may have slowed over time compared to the initial impact of the lockdown.

Comparing retail to other industries (as shown by the indname coefficients), most industries have a negative coefficient, indicating they fared worse than retail since these coefficients represent the difference from retail employment. However, Education & Healthcare has a positive coefficient, indicating this sector fared better than retail in terms of employment numbers during the period studied.

The high R-squared of the regression model implies that these variables explain a large proportion of the variance in employment, suggesting the model fits the data well.

	model2
Dependent Var.:	Employed
Constant	0.9412*** (0.0043)
post_lockdownTRUE	-0.0795*** (0.0090)
COVID_center	-2.43e-5. (1.43e-5)
indname = Agriculture&Mining	0.0114. (0.0059)
indname = Arts&Hospitality	-0.0716*** (0.0163)
indname = Construction	0.0049 (0.0051)
indname = Education&Healthcare	0.0286*** (0.0043)
indname = Finance&RealEstate	0.0422*** (0.0050)
indname = Information	0.0145*** (0.0041)
indname = Manufacturing	0.0218*** (0.0042)
indname = Military	-0.9228*** (0.0068)
indname = OtherServices	0.0033 (0.0068)
indname = Professional&Administrative	0.0181*** (0.0042)
indname = PublicAdministration	0.0501*** (0.0056)
indname = Transportation&Utilities	0.0102* (0.0041)
indname = WholesaleTrade	0.0303*** (0.0044)
post_lockdownTRUE x COVID_center	0.0002*** (3.75e-5)
S.E. type	Heteroskedast.-rob.
Observations	322
R2	0.98884
Adj. R2	0.98826

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Assumptions

- The industry categories accurately capture and group the relevant industries. This assumption seems reasonable as the industry categories are based on standard industry classifications.

- The data is representative of the overall employment trends in each industry.

- The model includes all relevant variables, and there are no omitted variable biases. It is possible that other relevant factors, such as industry-specific policies, regulations, or technological changes, could have influenced employment levels.

Conclusion

Retail employment saw a decrease following the COVID-19 lockdowns but fared relatively better than some industries like Agriculture & Mining, Arts & Hospitality, and Other Services.

The rate of decline in retail employment seemed to slow over time, indicating some recovery or adaptation. In contrast, sectors like Education & Healthcare experienced growth in employment, likely due to increased demand in healthcare services during the pandemic.

In conclusion, the retail sector was impacted by the lockdown but showed signs of resilience when compared to several other sectors. The analysis, while robust, is based on several assumptions, most of which are plausible given the context but may require further data or analysis to fully validate.

THANKS