Is Minimum Wage An Important Cause of Change for Unemployment and Income?

Nathan Perran, Yan Remes, Iavis Kopets, Jooyoung Park

6/7/2021

Introduction:

An increase in the minimum wage is often proposed towards the economic well-being within a society under the conventional perception that increased minimum wage helps households to be more financially stable. However, it is often argued that the proposed benefits of increasing minimum wage such as increasing personal income with same working hours do not outweigh the downside of employers considering laying off some of their employees to reduce some of their expenses and increasing overall unemployment. Considering these proposed consequences, would the effect of increasing minimum wage still help increasing household income while the employment rate is declining?

Through observing and processing historical data on minimum wage, household income and unemployment rate, our group aims to answer these questions. Our alternative hypothesis for the question is that unemployment rates will increase, but the total income of households will increase as lower paid workers gain more wage every working hour. Naturally, our null hypothesis is that there is no change in unemployment or household income for change in minimum wage.

Data:

Source:

We had two primary sources for our data. The first source where we got our minimum wage data came from a cleaned dataset on Kaggle which compiled minimum wage values in each state for each year as reported by the United States Department of Labor. The data set was last updated 5 months ago so it is up to date and relevant to our research. The second source we used was the FRED, St. Louis Fed, who extensively gathers economic data and conducts economic research. They get some of their data from the U.S. Bureau of Labor Statistics. We used the Fred website to strip the data we needed for each state, specifically to gather unemployment rates, the labor population, and median household income. Our data came from government/quasi-government websites that collect and report data. One can assume that the sources for the data are credible and portray accurate data for the basis of our research.

Measurements:

State (state) - list of all fifty states in the U.S.

Year (year) - years in which information was collected about different states starting from 1985 and until 2018.

Unemployment Rate (unem.rate) - shows the rate at which people were unemployed starting from 1985 and until 2018 in the different states in the U.S.

Minimum Wage (min.wage) - shows hourly minimum wage starting from 1985 and until 2018 in the different states in the U.S.

Household Income (house.income) - used as a measurement of economic livelihood. It indicates how much money the median household in the United States is making each year. For our purposes, median household income was chosen because it is the most available source of data for income in the United States, making our data more reliable. The real target population for the benefits of increasing minimum wage are low income individuals and households, and while we expect increases in minimum wage to be somewhat reflected in median household income, this measure isn't fully representative of our actual target demographic. However, because data on isn't readily available on a large scale for low income households, we chose to use median household income as the best estimate available to us.

Labor Populations (labor.pop) - working population in different states from 1985 to 2018.

Loading Data:

```
data <- read.csv("Compiled_Data.csv")
```

Preprocessing:

```
# create a sequence of elements to run in the following for loops, contents doesn't actually
# matter just requires the number of elements, but 1-50 represents index for troubleshooting
years <- seq(1985, 2018, 1)

num.states <- seq(1, 50, 1)

# set min and max acceptable proportions
min <- .9
max <- 1 / .9


comp <- data.frame(matrix(nrow = 0, ncol = 3))

# first for loop runs fifty times
for (i in years) {

  year.sub <- subset(data, year == i)

  for (j in num.states) {


    # because states don't need to be compared twice, starting at x + 1 means each state will
    # be compared to every state after it in the data frame
    for (k in num.states[j + 1: 50]) {

      # this if else statement takes the state at the index of the first for loop and compares
      # it to the state at the index of the second for loop, logical statements ensure that
      # proportion of each covariate for each state is within the specified min and max range,
```

```r
      # if they are comparable on each covariate it prints the name of each state to the console
      # I tried to create a string variable and this ifelse statement would add state pairs to
      # the variable and then that variable would contain all valid comparable states but it
      # kept redefining the variable instead of adding to it, figured printing to the console
      # and recording them manually would be fine


      ifelse(year.sub$unem.rate[j] / year.sub$unem.rate[k] >= min & year.sub$unem.rate[j] /
                year.sub$unem.rate[k] <= max
            & year.sub$min.wage[j] / year.sub$min.wage[k] >= min & year.sub$min.wage[j] /
                year.sub$min.wage[k] <= max
            & year.sub$house.income[j] / year.sub$house.income[k] >= min & year.sub$house.income[j] /
                year.sub$house.income[k] <= max
            & year.sub$labor.pop[j] / year.sub$labor.pop[k] >= min & year.sub$labor.pop[j] /
                year.sub$labor.pop[k] <= max,
            comp <- rbind(comp, c(year.sub$year[1], year.sub$state[j], year.sub$state[k])), NA)
    }
  }
}

# rename columns of new data frame
names(comp) <- c("year", "state1", "state2")

# remove years where the US was going through an economic recession
comp <- subset(comp, year != 1990 & year != 1991 & year != 2001 & year != 2008 & year != 2009)
rownames(comp) <- NULL

# create a new data frame for further selective sample
comparable <- data.frame(matrix(nrow = 0, ncol = 3))
comp$year <- as.integer(comp$year)

for (i in 1:nrow(comp)) {

  # subset original data frame to only include data for current pair of matched states
  # in the year of being matched
  possible.match <- subset(data, (year == comp$year[i] | year == comp$year[i] + 1) &
                             (state == comp$state1[i] | state == comp$state2[i]))

  # make sure that one state had an increase in minimum wage and the other state had
  # no change
  ifelse(((possible.match$min.wage[2] - possible.match$min.wage[1]) > 0 &
            possible.match$min.wage[4] == possible.match$min.wage[3]) |
          ((possible.match$min.wage[4] - possible.match$min.wage[3]) > 0 &
             possible.match$min.wage[2] == possible.match$min.wage[1]),
        comparable <- rbind(comparable, c(possible.match$year[1], possible.match$state[1],
                                          possible.match$state[3])), NA)

}
colnames(comparable) <- c("year", "state1", "state2")


for (i in 1:nrow(comparable)) {
```

```r
  # subset original data frame to only include data for current pair of matched states
  # in the year of being matched
  matched <- subset(data, (year == as.integer(comparable$year[i]) | year ==
                          as.integer(comparable$year[i]) + 1) &
                    (state == comparable$state1[i] | state == comparable$state2[i]))

  # reorganize data frame so state with a change is state1 in the data frame
  comparable$change.min.wage[i] <- max(matched$min.wage[2] - matched$min.wage[1],
                                       matched$min.wage[4] - matched$min.wage[3])

  place.holder <- matched

  comparable$state1[i] <- ifelse((matched$min.wage[2] - matched$min.wage[1] == 0),
                                 place.holder$state[3], place.holder$state[1])

  comparable$state2[i] <- ifelse((matched$min.wage[2] - matched$min.wage[1] == 0),
                                 place.holder$state[1], place.holder$state[3])
}

# create a new data frame for further selective sample
comparable2 <- data.frame(matrix(nrow = 0, ncol = 4))
comparable$year <- as.integer(comparable$year)

for (i in 1:nrow(comparable)) {
  # subset original data frame to only include data for current pair of matched states
  # in the year of being matched
  possible.match <- subset(data, (year == comparable$year[i] | year == comparable$year[i] + 1) &
                           (state == comparable$state1[i] | state == comparable$state2[i]))

  # make sure that each of our matched variables have the same slope between states
  # trying to ensure that they have the same trajectory and they don't just happen to
  # have the same value while one is decreasing and the other is rising, because that
  # isn't really the same
  ifelse(((possible.match$labor.pop[2] - possible.match$labor.pop[1]) > 0 &
           possible.match$labor.pop[4] > possible.match$labor.pop[3]) |
         ((possible.match$labor.pop[4] - possible.match$labor.pop[3]) < 0 &
            possible.match$labor.pop[2] < possible.match$labor.pop[1]) &
         ((possible.match$unem.rate[2] - possible.match$unem.rate[1]) > 0 &
            possible.match$unem.rate[4] > possible.match$unem.rate[3]) |
         ((possible.match$unem.rate[4] - possible.match$unem.rate[3]) < 0 &
            possible.match$unem.rate[2] < possible.match$unem.rate[1]) &
         ((possible.match$house.income[2] - possible.match$house.income[1]) > 0 &
            possible.match$house.income[4] > possible.match$house.income[3]) |
         ((possible.match$house.income[4] - possible.match$house.income[3]) < 0 &
            possible.match$house.income[2] < possible.match$house.income[1]),
       comparable2 <- rbind(comparable2, c(possible.match$year[1], possible.match$state[1],
                            possible.match$state[3], round(comparable$change.min.wage[i], 2))), NA)
}


colnames(comparable2) <- c("year", "state1", "state2", "change.min.wage")

# make sure the year is before 2016 because our final measurements are taken 2 years
```

```r
# after and 2018 is the last year for our data
comparable2 <- subset(comparable2, year <= 2016)

# create a data frame for final measurements
final.after <- data.frame(matrix(nrow = nrow(comparable2), ncol = 10))
colnames(final.after) <- c("year", "state1", "state1.min.wage", "state1.unem.rate",
                           "state1.house.income", "state2", "state2.min.wage",
                           "state2.unem.rate", "state2.house.income", "change.min.wage")

final.after$year <- as.integer(comparable2$year) + 2
final.after$state1 <- comparable2$state1
final.after$state2 <- comparable2$state2

# this for loop fills out the final data frame with values from the comparable one
for (i in 1:nrow(final.after)) {
  final.after$state1.house.income[i] <- data$house.income[data$year == final.after$year[i] &
                                            data$state == final.after$state1[i]]

  final.after$state1.min.wage[i] <- data$min.wage[data$year == final.after$year[i] &
                                        data$state == final.after$state1[i]]

  final.after$state1.unem.rate[i] <- data$unem.rate[data$year == final.after$year[i] &
                                        data$state == final.after$state1[i]]

  final.after$state2.house.income[i] <- data$house.income[data$year == final.after$year[i] &
                                            data$state == final.after$state2[i]]

  final.after$state2.min.wage[i] <- data$min.wage[data$year == final.after$year[i] &
                                        data$state == final.after$state2[i]]

  final.after$state2.unem.rate[i] <- data$unem.rate[data$year == final.after$year[i] &
                                        data$state == final.after$state2[i]]

  final.after$change.min.wage[i] <- comparable2$change.min.wage[i]

}

# create a data frame for initial measurements
final.before <- data.frame(matrix(nrow = nrow(comparable2), ncol = 9))
colnames(final.before) <- c("year", "state1", "state1.min.wage", "state1.unem.rate",
                            "state1.house.income", "state2", "state2.min.wage",
                            "state2.unem.rate", "state2.house.income")

final.before$year <- as.integer(comparable2$year)
final.before$state1 <- comparable2$state1
final.before$state2 <- comparable2$state2

# this for loop fills out the initial data frame with values from the comparable one
for (i in 1:nrow(final.before)) {
  final.before$state1.house.income[i] <- data$house.income[data$year == final.before$year[i] &
                                            data$state == final.before$state1[i]]

  final.before$state1.min.wage[i] <- data$min.wage[data$year == final.before$year[i] &
```

```
                                                    data$state == final.before$state1[i]]

  final.before$state1.unem.rate[i] <- data$unem.rate[data$year == final.before$year[i] &
                                                data$state == final.before$state1[i]]

  final.before$state2.house.income[i] <- data$house.income[data$year == final.before$year[i] &
                                                      data$state == final.before$state2[i]]

  final.before$state2.min.wage[i] <- data$min.wage[data$year == final.before$year[i] &
                                                data$state == final.before$state2[i]]

  final.before$state2.unem.rate[i] <- data$unem.rate[data$year == final.before$year[i] &
                                                data$state == final.before$state2[i]]


}


dataBefore <- final.before
dataAfter <- final.after

# ATE is misnamed but calculates the differences between the matched pairs
ATEstate1 <- dataAfter[1:nrow(dataAfter), 4:5] - dataBefore[1:nrow(dataBefore), 4:5]
ATEstate2 <- dataAfter[1:nrow(dataAfter), 8:9] - dataBefore[1:nrow(dataBefore), 8:9]
combineState1 <- cbind(dataAfter[ , 1:3], ATEstate1)
combineAll <- cbind(combineState1, dataAfter$state2, ATEstate2)
colnames(combineAll)[6] <- "state2"

# difference in difference
difference <- combineAll[ , 7:8] - combineAll[ , 4:5]
colnames(difference) <- c("difference.unem.rate", "difference.house.income")
difference$change.min.wage <- as.numeric(final.after$change.min.wage)
```

Dependent Variable Graphs:

```
# find data points for used states in respective years and give them an index that
# signifies they belong to a particular data point in the final graphs
dependent.var <- data.frame(matrix(nrow = 0, ncol = 5))
for (i in 1:nrow(final.before)) {
  state1 <- subset(data, select = (c("year", "state", "house.income", "unem.rate")),
                   subset = (state == final.before$state1[i] & year >= final.before$year[i] &
                              year <= (final.before$year[i] + 2)))
  state1$index <- i
  dependent.var <- rbind(dependent.var, state1)

  state2 <- subset(data, select = (c("year", "state", "house.income", "unem.rate")),
                   subset = (state == final.before$state2[i] & year >= final.before$year[i] &
                              year <= (final.before$year[i] + 2)))
  state2$index <- i
  dependent.var <- rbind(dependent.var, state2)
}
```
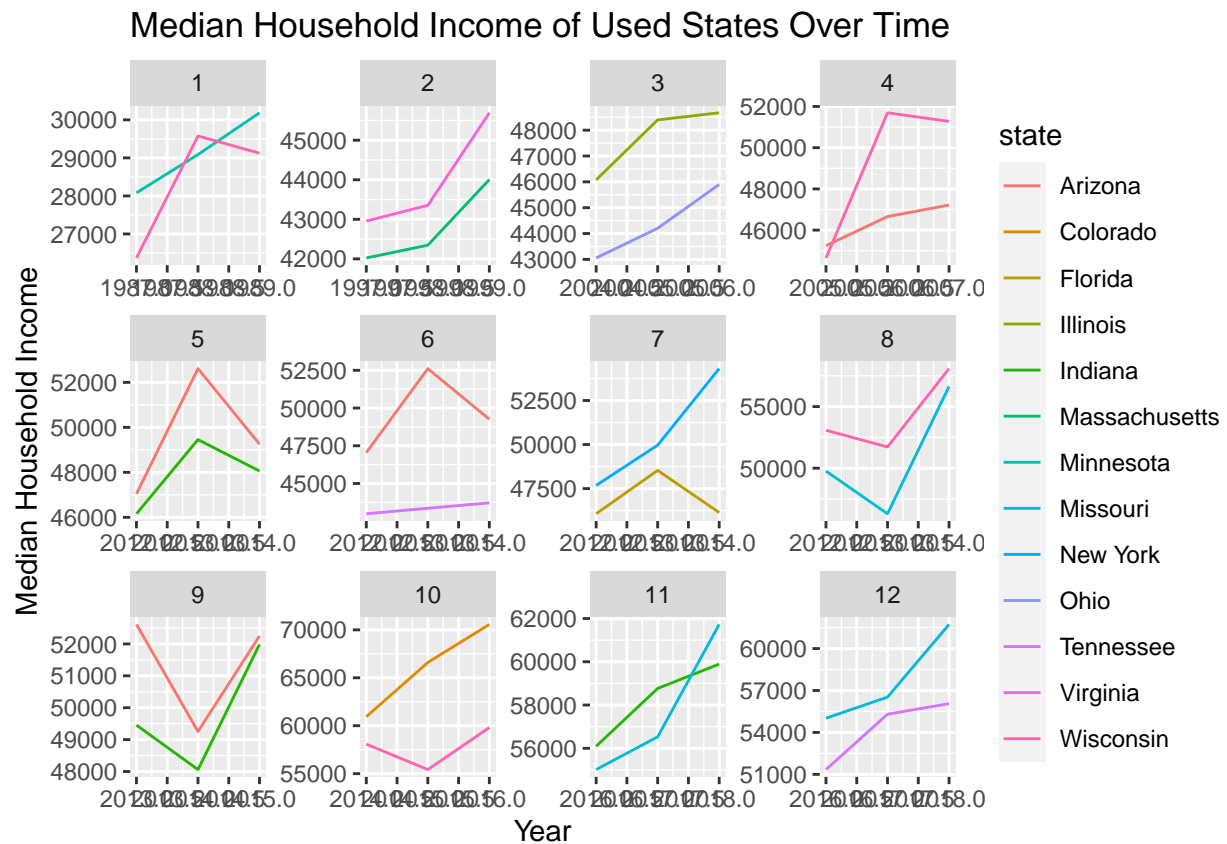
```
rownames(dependent.var) <- NULL

# plot dependent variables
ggplot(dependent.var, aes(x = year, y = house.income, color = state)) +
  geom_line() +
  labs(x = "Year", y = "Median Household Income", title =
         "Median Household Income of Used States Over Time") +
  facet_wrap(~ index, scales = "free")
```
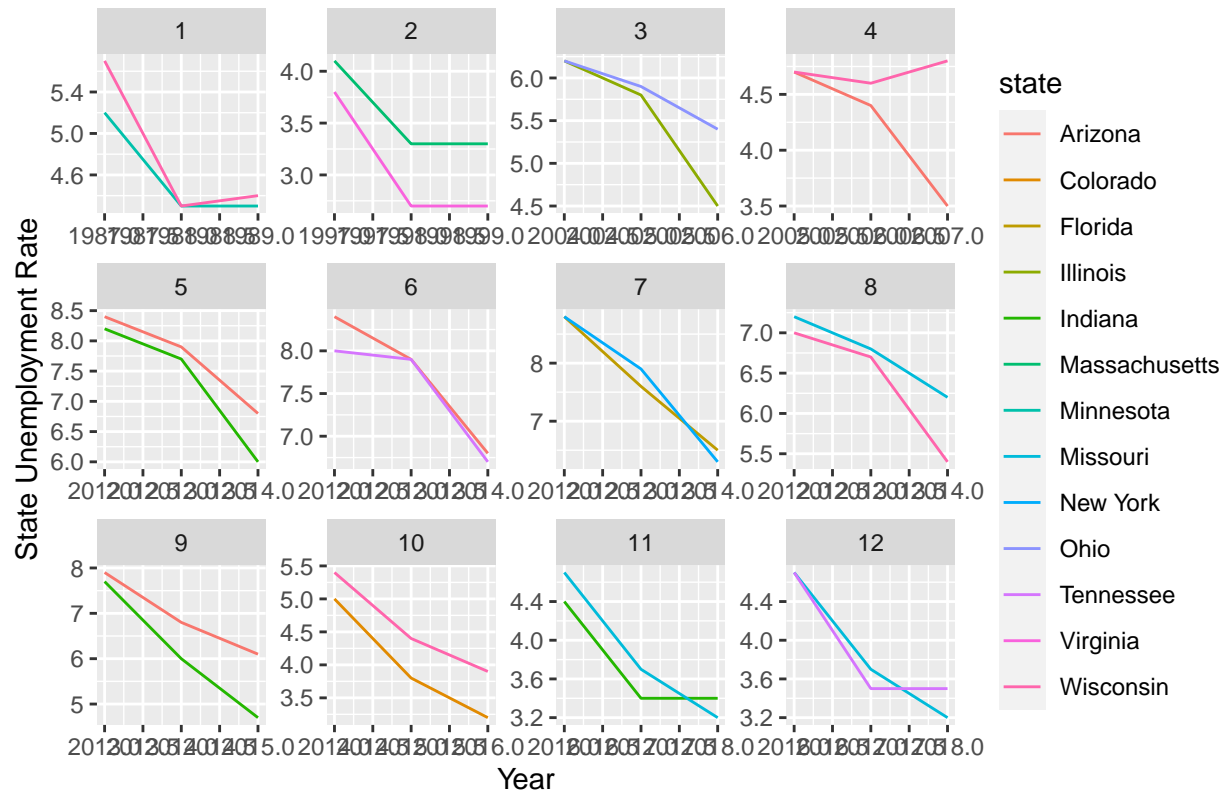

Median Household Income of Used States Over Time

```
ggplot(dependent.var, aes(x = year, y = unem.rate, color = state)) +
  geom_line() +
  labs(x = "Year", y = "State Unemployment Rate", title =
         "Unemployment Rate of Used States Over Time") +
  facet_wrap(~ index, scales = "free")
```

Unemployment Rate of Used States Over Time


Results:

```r
#plot data
dif.unem <- ggplot(difference, aes(x = change.min.wage, y = difference.unem.rate)) +
  geom_point() +
  geom_smooth(method = "lm") +
  labs(x = "Change in Minimum Wage", y = "Change in Unemployment Rate", title =
       "Difference in Differences for Unemployment Rate")

dif.income <- ggplot(difference, aes(x = change.min.wage, y = difference.house.income)) +
  geom_point() +
  geom_smooth(method = "lm") +
  labs(x = "Change in Minimum Wage", y = "Change in Median Household Income", title =
       "Difference in Differences for Median Household Income")

dif.unem
```

```
## `geom_smooth()` using formula 'y ~ x'
```
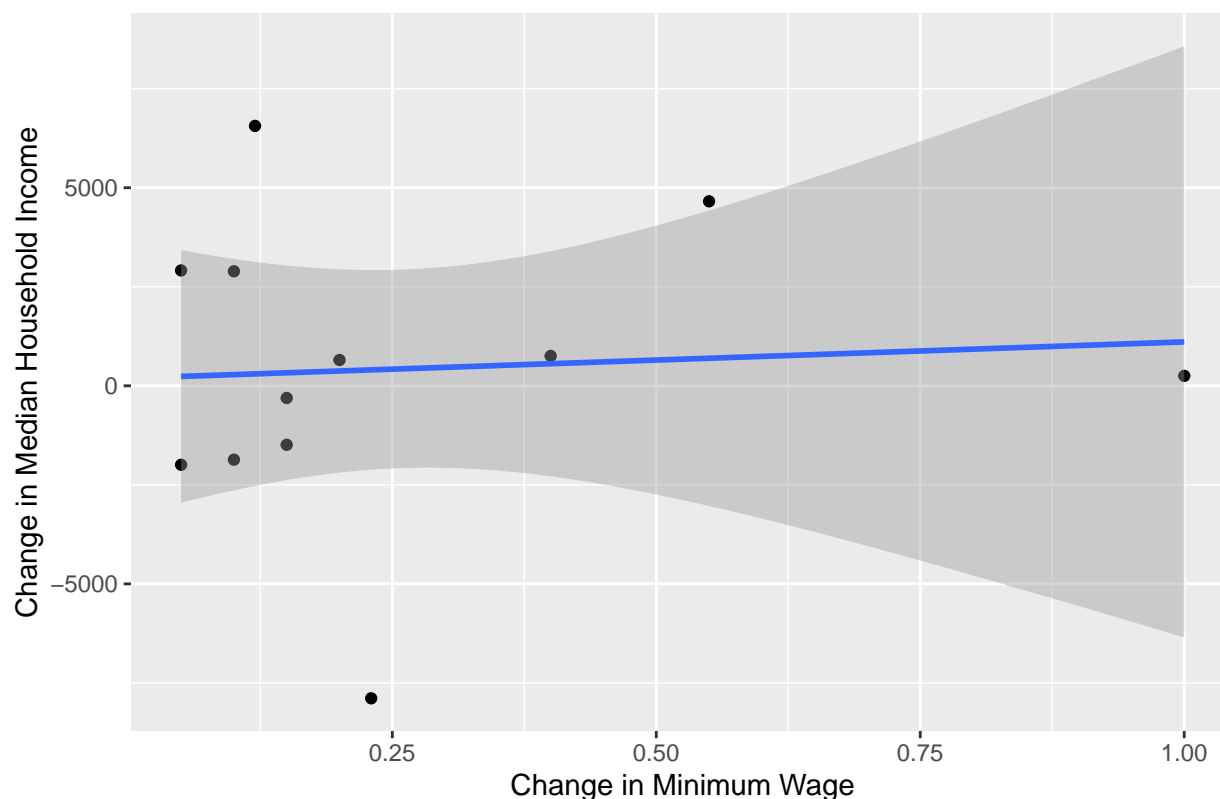
Difference in Differences for Unemployment Rate
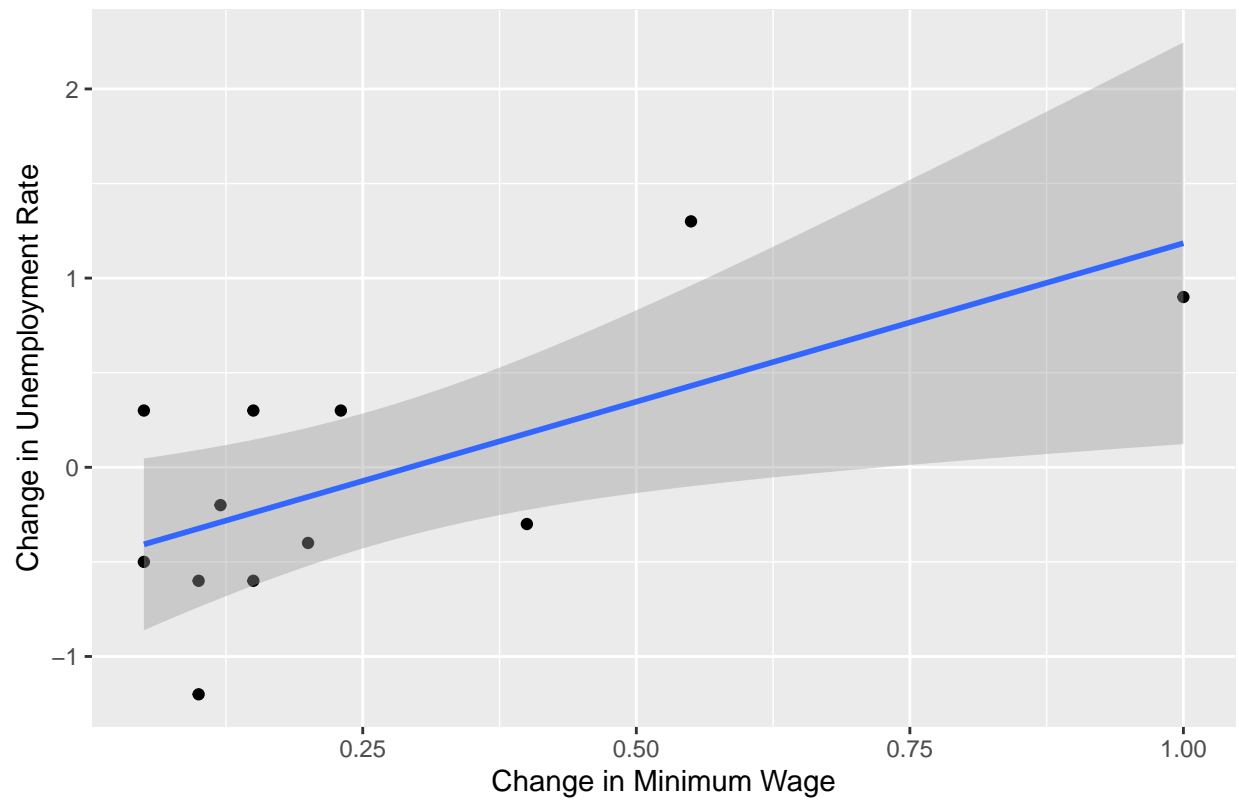

```
dif.income
```

```
## 'geom_smooth()' using formula 'y ~ x'
```

Difference in Differences for Median Household Income

Conclusion:

The first graph above depicts how unemployment changes in the years following a change in a state's minimum wage. The graph shows that a very minor change in minimum wage at around 10 cents actually causes a decrease in unemployment in the following years, with larger changes to minimum wage at a dollar causing a 1% increase in unemployment. The end of the trend line is consistent with our hypothesis that increases in minimum wage cause an increase in unemployment, but we were surprised by the decrease in unemployment at lower increases in minimum wage.

The second graph depicts an almost flat trend line around 0 for changes in median household income as a result of changes in minimum wage. This goes against our original hypothesis that household income would increase as a result of increasing minimum wage, but we should reiterate that the important demographic for increasing minimum wage is low income families and it is possible that this effect is still occurring, but simply not captured by the measure of median household income. Therefore, this data is certainly not evidence that increasing minimum wage increases income of poor households, but it is also questionable as evidence against the claim that increasing the minimum wage increases income for poor households. The biggest threat to inference in our analysis is whether or not our matched pairs of states are actually reasonable substitutes for each other and if we could assume that the future values of unemployment and household income would be the same between these two states if one of them hadn't received the treatment of a change in minimum wage. We did our best to make states as comparable as possible, but without the ability to randomize which states get the treatment and which ones don't, it's impossible to account for everything.

In our process for finding pairs of states that were comparable enough for one to provide the counterfactual for the other receiving the treatment of a change in minimum wage, we made the arbitrary decision to say that states were comparable if one states value for a variable divided by the other states value was greater than

or equal to 0.9. This ensured that states had similar starting minimum wages, median household incomes, unemployment rates, and labor populations. When we increased this threshold, our resulting regression lines remained similar until there weren't enough data points to make any reasonable inference. When decreasing this threshold, the regression lines again remain fairly similar until about 0.8, at which point we don't think the states are really comparable and the regression lines are no longer meaningful.

Another threat to inference is that changes in economic policy often have delayed results as the economy is a deeply interconnected construction where a small change in one area ripples out and can have profound impacts in other areas. The point is that the manifestation of these effects take time to appear, and we tried to account for this by taking our secondary observations for values of variables a full year after the initial change was made. This introduces the issue that other changes can occur in that year between observations such as additional changes in minimum wage in either the control or treatment state, but we thought that those secondary changes in minimum wage would themselves have delayed effects and that overall this choice would increase internal validity.

The final threat to inference is that it is possible that states are self-selecting into the treatment group and that there is a systematic difference between states that choose to increase minimum wage and those that don't. We considered dealing with this by only including states in the treatment group who increased minimum wage because of a federal increase in minimum wage, but this doesn't solve the issue of self-selection because then states would be putting themselves into the control group by having a minimum wage higher than the federal one and not needing to increase their own to meet federal requirements.