Econometrics Final Project: The Effects of Minimum Wage on Unemployment

Ted Johnson, Alicia Xu, Shalem Sumanthiran

12/16/2023

Introduction

For our project, we wanted to explore whether minimum wage laws significantly impact unemployment rates. This topic could soon become relevant to us as New York City is increasing its hourly minimum wage to $16 at the start of 2024, and in the midst of what many people fear is an economic recession, we wanted to know whether legislation like this actually has any impact. While there are many areas that minimum wage laws impact, we chose to focus on unemployment rates as they are directly related and simple to measure. In labor economics literature, the effect of minimum wage is widely disputed. Some argue that raising the minimum wage does not increase unemployment (Card and Krueger), while others argue that the adverse effects of minimum wage are hidden in job growth rates (Meer and West) and long-term unemployment (Clemens and Wither). In our project, we looked at long-term effects of minimum wage laws across different states, with data spanning from 1976 to 2022. After estimating multiple models and comparing their fits, we concluded that there is a positive relationship between minimum wage and unemployment.

Survey of the Literature

One of the most cited econometric studies on the effects of minimum wage was published by David Card and Alan Krueger in 1994. They observed a natural experiment in regions of New Jersey and eastern Pennsylvania in 1992, when New Jersey raised its minimum wage but Pennsylvania did not. They observed the impact of the legislation specifically on the fast food industry, surveying about 400 fast food chain restaurants. They chose to focus on fast food employment because they felt it was representative of minimum wage employment, and the data on employment would be easiest to access. They collected the data through two waves of interviews, one before and one after the minimum wage increase.

Card and Krueger's initial analysis of their survey results looked at the average of full time employment and wages in the first wave versus the second, segmented by New Jersey and Pennsylvania stores. They then subtracted the two averages for each metric to find the difference between the two states. This method would be most similar to using entity-demeaned regression to estimate panel data; however, Card and Krueger did not treat it as panel data in the initial analysis. The initial analysis found that while the average wage increased in New Jersey, full time employment actually increased, which would be contrary to most economic reasoning.

Suspecting that there were omitted variables such as franchise management, Card and Krueger utilized fixed effect regression to control for additional variables. In their regression, they made New Jersey as a dummy variable, added control variables such as franchise ownership, and entity fixed effects for particular regions of New Jersey/Pennsylvania. Finally, they used heteroskedastic-robust standard errors to account for differences in variance between small stores and large stores. They found that adding the controls and fixed effects improved the fit of their models, and the model estimate provided statistically significant evidence that the increase in minimum wage did positively impact employment.

Because their analysis contradicted economic theory, Card and Krueger tried to eliminate outliers and omitted variables by removing specific regions of New Jersey, factoring in part-time employment and store openings, and comparing wage changes within New Jersey alone. In these analyses, they continued to find evidence that the increase in minimum wage actually increased employment. They propose a number of economic models to explain this contradictory effect, but their analysis does not fully support any of these explanations. Their findings are somewhat inconclusive, but they definitively refute the idea that increasing minimum wage immediately lowers employment.

In 2013, Jonathan Meer and Jeremy West analyzed the effects of minimum wage on job growth rates using broader panel data on all states. Their initial findings, using a time and state fixed effects regression, aligned with those of Card and Krueger, suggesting that minimum wage does not increase unemployment in the short run. However, Meer and West focus their model estimation more on the variables that impact the relationship between unemployment and minimum wage. Using a Monte Carlo simulation, they demonstrate that a time trend in regression actually biases the coefficient of minimum wage towards zero. They also use state real GDP per capita as an important control variable to reflect state-level fluctuations in the business cycle. Instead of using a time fixed effect, they estimate a distributed lag time series model to demonstrate the long-term effects of minimum wage on employment. They conclude that the negative effects of minimum wage are statistically significant in the long term, with decreases in employment occurring after about 2-3 years. This complicates the relationship suggested in Card and Krueger's analysis, as they only looked at the effects of the legislation after less than one year. Furthermore, Meer and West analyze a much larger scope of data, which makes their findings more broadly applicable but also creates more possibilities for omitted variable bias.

Finally, Jeffrey Clemens and Michael Wither's 2014 paper on the effects of minimum wage during a recession demonstrates the difference in effects of minimum wage on different population groups. Clemens and Wither focus on a specific period of minimum wage increase from 2007 to 2009 during the Great Recession. Their regression utilizes state and time fixed effects, as well as a time trend term. Furthermore, they segment their analysis demographically, focusing on two main segments of minimum wage workers: teenagers and low-income adults. Their analysis finds that the increase in minimum wage significantly lowered the likelihood of low-income workers moving into the middle class in the long term.

Based on the existing literature, there is conflicting evidence about how minimum wage legislation impacts unemployment. Its effects are clearly highly variable across different time periods, geographic regions, and demographics. For these reasons, using panel data with fixed effects and control variables is very common, and it will be important to consider many potential omitted variables. The relationship between the two variables over time is also highly disputed, so using data with a large time horizon will also have an impact on our results.

Model

The primary model we are using is based on panel data using time and state fixed effects and control variables. This model is commonly used in previous literature about minimum wage and unemployment, and it will also allow us to control for state-specific minimum wage laws as well as macroeconomic fluctuations in business cycles. The model follows the general form:

$$UN_{it} = \beta_0 + \beta_1 Mwage_{it} + \beta_2 Z_{it} + \alpha_i + \delta_t + u_{it}$$

Where UN represents unemployment, Mwage represents minimum wage, and Z represents any control variable (which will vary in number). We will fit both state and time fixed effects as well as control variables to see what results in the best fit regression.

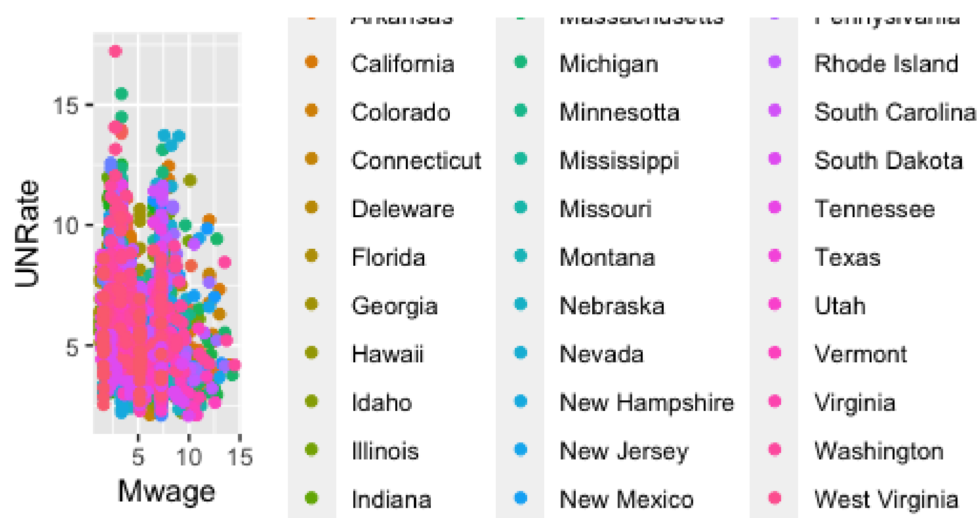
Data

The dataset we are using is from FRED, and it contains the Average Unemployment rate per year from 1976 to 2022 for each state, and the minimum wage for each state from 1976 to 2022. Alabama, Louisiana, Mississippi, South Carolina, and Tennessee do not have a state minimum wage and instead adhere to the federal minimum wage; this is reflected in the data. The dataset also includes the variables Real GDP (from 2017 to 2022), education level (from 2006 to 2022), and income level (from 1984 to 2022). Because these variables are not available for all time periods, we do need to filter the data set to certain years when including them into our regression model.

Empirical Application

First, we plotted the data on a scatter plot. Looking at the graph, there appears to be a slight negative relationship between the two variables, given the empty triangle in the top right corner. It is clear that there is heteroskedasticity in the data, as we see larger variance for lower minimum wages. There do not appear to be any major outliers.

```
###Importing the data
setwd("~/")
library(readr)
library(ggplot2)
library(readxl)
library(dplyr)
mydata <- read_excel(path = "Downloads/Econ Project Data.xlsx")
mydata2 <- filter(mydata, Year > "2016")
head(mydata2)

# A tibble: 6 × 8
  State    Year stateid UNRate Mwage    rgdp educL incomeL
  <chr>   <dbl>   <dbl>  <dbl> <dbl>   <dbl> <dbl>   <dbl>
1 Alabama  2017       1   4.51  7.25 216616.  25.5   59980
2 Alabama  2018       1   3.94  7.25 220809.  25.5   57720
3 Alabama  2019       1   3.18  7.25 224945.  26.3   64010
4 Alabama  2020       1   6.42  7.25 222081.  27.8   61650
5 Alabama  2021       1   3.37  7.25 231893.  27.4   61390
6 Alabama  2022       1   2.58  7.25 235807.  28.8   59910

head(mydata)

# A tibble: 6 × 8
  State    Year stateid UNRate Mwage  rgdp educL incomeL
  <chr>   <dbl>   <dbl>  <dbl> <dbl> <dbl> <dbl>   <dbl>
1 Alabama  1976       1   6.7   2.3     NA    NA      NA
2 Alabama  1977       1   7.15  2.3     NA    NA      NA
3 Alabama  1978       1   6.41  2.3     NA    NA      NA
4 Alabama  1979       1   7.22  2.9     NA    NA      NA
5 Alabama  1980       1   8.82  3.1     NA    NA      NA
6 Alabama  1981       1  10.7   3.35    NA    NA      NA

###Plotting the Data
ggplot(mydata, aes(Mwage,UNRate, color = State)) +
  geom_point()
```



We attempted a log transformation of the data to see if this would help visualize any trend.

```
###Log Transformation
#mydata <- mutate(mydata, LMwage = log(Mwage))
mydata <- mutate(mydata, LUNRate = log(UNRate))
ggplot(mydata, aes(Mwage,LUNRate, color = State)) +
  geom_point()
```

After performing the log transformation, there is minimal difference in the data visually. For this reason, we elected to proceed with the non-transformed data for our analysis. We proceeded to fit a variety of models to the data in order to assess what would best fit the data and explain the relationship between unemployment and minimum wages without bias.

Pooled Regression

We started by defining the data set as a panel with indexes state and year, and estimating the pooled regression model. We used heteroskedastic adjusted errors for the panel in order to account for differences in variance across states.

```
###Defining Data Set as Panel
library(plm)
mydata.pd <- pdata.frame(mydata, index=c("stateid", "Year"),
                         drop.index = TRUE, row.names = TRUE)
###Running Pooled Regression
mydata.pool <- plm(UNRate ~ Mwage, data = mydata.pd, model = "pooling")
summary(mydata.pool, vcov = vcovHC)

Pooling Model

Note: Coefficient variance-covariance matrix supplied: vcovHC

Call:
plm(formula = UNRate ~ Mwage, data = mydata.pd, model = "pooling")
```

```
Balanced Panel: n = 50, T = 47, N = 2350

Residuals:
    Min.  1st Qu.   Median  3rd Qu.     Max. 
-3.87485 -1.48988 -0.37383  1.14144 11.06321 

Coefficients:
             Estimate Std. Error t-value              Pr(>|t|)    
(Intercept)  6.568620   0.226997  28.937 < 0.00000000000000022 ***
Mwage       -0.144907   0.025881  -5.599         0.00000002408 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    10251
Residual Sum of Squares: 9959.8
R-Squared:      0.028446
Adj. R-Squared: 0.028032
F-statistic: 31.3484 on 1 and 49 DF, p-value: 0.00000096476
```

Looking at the estimate for pooled regression, there appears to be a negative relationship between Minimum Wage and Unemployment Rate. The coefficient of minimum wage is statistically significant at the 5% level with a very low t-value of -5.599. This would be consistent with the findings in Card and Krueger's analysis, as increases in minimum wage would decrease the unemployment rate. For a $1 increase in minimum wage, we would expect a concurrent fall in the unemployment rate by .14%, all other factors held equal.

Running State and Time Fixed Effects

Finally, we wanted to factor in the control variables we had available (income level, education, and state GDP) as well as state and time fixed effects in order to see if the initial pooled model was affected by bias.

```r
# (i) Time and State Fixed Effects Only
mymodel.fe <- plm(UNRate ~ Mwage, data = mydata.pd,
                  model = "within", effect = "twoways")
summary(mymodel.fe, vcov = vcovHC)
```

```
Twoways effects Within Model

Note: Coefficient variance-covariance matrix supplied: vcovHC

Call:
plm(formula = UNRate ~ Mwage, data = mydata.pd, effect = "twoways", 
    model = "within")

Balanced Panel: n = 50, T = 47, N = 2350

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max. 
```

```
-3.377296 -0.621224 -0.033783  0.557988  5.963420

Coefficients:
      Estimate Std. Error t-value Pr(>|t|)
Mwage 0.052525   0.059673  0.8802   0.3788

Total Sum of Squares:    2530.7
Residual Sum of Squares: 2527.5
R-Squared:      0.001247
Adj. R-Squared: -0.04131
F-statistic: 0.774794 on 1 and 49 DF, p-value: 0.38304
```

```
fixef(mymodel.fe)
```

```
     1      2      3      4      5      6      7      8      9     10     11
7.9530 8.7726 7.2893 7.2982 8.2767 6.3984 6.4752 6.4576 7.0720 7.0708 5.8704
    12     13     14     15     16     17     18     19     20     21     22
6.8625 7.9384 7.0635 5.6194 5.7527 8.1162 6.6984 6.3185 6.5857 8.8339 5.8516
    23     24     25     26     27     28     29     30     31     32     33
8.4546 6.8611 7.6565 6.6938 4.5555 7.7879 5.3302 7.7903 7.5741 6.9019 4.8475
    34     35     36     37     38     39     40     41     42     43     44
7.7161 6.0343 7.8846 7.3247 7.4926 7.4489 7.4216 7.3325 7.1150 5.7875 5.5607
    45     46     47     48     49     50
5.7766 7.7435 9.0714 6.4066 4.6965 6.0463
```

```
# (ii) Time and State Fixed Effects + education
mymodel.incfe <- plm(UNRate ~ Mwage+educL, data = mydata.pd,
                     model = "within", effect = "twoways")
summary(mymodel.incfe, vcov = vcovHC)
```

```
Twoways effects Within Model

Note: Coefficient variance-covariance matrix supplied: vcovHC

Call:
plm(formula = UNRate ~ Mwage + educL, data = mydata.pd, effect = "twoways",
    model = "within")

Balanced Panel: n = 50, T = 17, N = 850

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-2.723365 -0.547762 -0.015875  0.511682  5.382786

Coefficients:
       Estimate Std. Error t-value Pr(>|t|)
Mwage  0.048774   0.072616  0.6717  0.50199
educL -0.166780   0.078258 -2.1312  0.03339 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Total Sum of Squares:    586.7
Residual Sum of Squares: 577.97
R-Squared:      0.014886
Adj. R-Squared: -0.069516
F-statistic: 2.3021 on 2 and 49 DF, p-value: 0.11077
```

```r
# (iii) Time and State Fixed Effects + income + educ
mymodel.incedufe <- plm(UNRate ~ Mwage+educL+incomeL, data = mydata.pd,
                  model = "within", effect = "twoways")
summary(mymodel.incedufe, vcov = vcovHC)
```

```
Twoways effects Within Model

Note: Coefficient variance-covariance matrix supplied: vcovHC

Call:
plm(formula = UNRate ~ Mwage + educL + incomeL, data = mydata.pd,
    effect = "twoways", model = "within")

Balanced Panel: n = 50, T = 17, N = 850

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-2.467762 -0.550783 -0.025868  0.514036  5.381725

Coefficients:
            Estimate   Std. Error t-value Pr(>|t|)
Mwage    0.069774143  0.074536868  0.9361 0.349510
educL   -0.133737030  0.070819585 -1.8884 0.059340 .
incomeL -0.000033462  0.000012128 -2.7590 0.005933 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    586.7
Residual Sum of Squares: 567.6
R-Squared:      0.032566
Adj. R-Squared: -0.051667
F-statistic: 2.96023 on 3 and 49 DF, p-value: 0.041247
```

```r
# (iv) Time and State Fixed Effects + income + educ + gdp
mymodel.incedugdpfe <- plm(UNRate ~ Mwage+educL+incomeL+rgdp, data =
mydata.pd,
                  model = "within", effect = "twoways")
summary(mymodel.incedugdpfe, vcov = vcovHC)
```

```
Twoways effects Within Model

Note: Coefficient variance-covariance matrix supplied: vcovHC
```

```
Call:
plm(formula = UNRate ~ Mwage + educL + incomeL + rgdp, data = mydata.pd,
    effect = "twoways", model = "within")

Balanced Panel: n = 50, T = 6, N = 300

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-2.1817643 -0.3417076  0.0014785  0.3025188  4.3276269

Coefficients:
             Estimate    Std. Error t-value Pr(>|t|)
Mwage    0.1471347622  0.0608585018  2.4177  0.01636 *
educL    0.1750268347  0.0979254129  1.7873  0.07514 .
incomeL -0.0000122685  0.0000184752 -0.6641  0.50729
rgdp    -0.0000001891  0.0000015099 -0.1252  0.90044
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    132.44
Residual Sum of Squares: 126.94
R-Squared:      0.04151
Adj. R-Squared: -0.18916
F-statistic: 2.99929 on 4 and 49 DF, p-value: 0.027199
```

```r
# (v) Time Fixed Effects + income + educ + gdp
mymodel.incedugdpfe <- plm(UNRate ~ Mwage+educL+incomeL+rgdp, data =
mydata.pd,
                model = "within", effect = "time")
summary(mymodel.incedugdpfe, vcov = vcovHC)
```

```
Oneway (time) effect Within Model

Note: Coefficient variance-covariance matrix supplied: vcovHC

Call:
plm(formula = UNRate ~ Mwage + educL + incomeL + rgdp, data = mydata.pd,
    effect = "time", model = "within")

Balanced Panel: n = 50, T = 6, N = 300

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-3.182440 -0.600978 -0.030319  0.517519  6.321892

Coefficients:
              Estimate     Std. Error t-value   Pr(>|t|)
Mwage    0.21991874255  0.05560477233  3.9550 0.00009625 ***
educL   -0.03534886158  0.04003087291 -0.8830   0.377946
```

```
incomeL -0.00001441821  0.00001628373 -0.8854   0.376655
rgdp     0.00000055896  0.00000016902  3.3071   0.001061 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    375.49
Residual Sum of Squares: 304.46
R-Squared:      0.18915
Adj. R-Squared: 0.16399
F-statistic: 11.1889 on 4 and 49 DF, p-value: 0.0000015832

# (vi) State Fixed Effects + income + educ + gdp
mymodel.incedugdpfe <- plm(UNRate ~ Mwage+educL+incomeL+rgdp, data =
mydata.pd,
                  model = "within", effect = "individual")
summary(mymodel.incedugdpfe, vcov = vcovHC)

Oneway (individual) effect Within Model

Note: Coefficient variance-covariance matrix supplied: vcovHC

Call:
plm(formula = UNRate ~ Mwage + educL + incomeL + rgdp, data = mydata.pd,
    effect = "individual", model = "within")

Balanced Panel: n = 50, T = 6, N = 300

Residuals:
    Min.  1st Qu.   Median  3rd Qu.     Max.
-3.09951 -0.88221 -0.19035  0.66924  7.12158

Coefficients:
             Estimate    Std. Error t-value     Pr(>|t|)
Mwage   -0.1183507186  0.1151111874 -1.0281       0.3049
educL    0.4345284201  0.0798428922  5.4423 0.0000001271 ***
incomeL  0.0000294880  0.0000227704  1.2950       0.1965
rgdp    -0.0000125930  0.0000068194 -1.8466       0.0660 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    699.28
Residual Sum of Squares: 596.79
R-Squared:      0.14656
Adj. R-Squared: -0.037307
F-statistic: 12.1854 on 4 and 49 DF, p-value: 0.00000059433
```

Based on the t-tests, the coefficient of minimum wage is statistically significant at a 5% level in Model (iv) and at a 1% level in Model (v). Because adding more variables resulted in the coefficient for Mwage increasing from negative to positive, this would suggest that there was omitted variable bias in our original pooled regression model. The decrease in unemployment was probably attributable to other variables, like income and state-specific effects. Furthermore, the F-stat p-value decreases as we add regressors, which suggests that at least one of the additional variables does have nonzero effects on unemployment. Out of these models, I would select model (v) based on its very low F-stat p-value and relatively high R^2 value.

If we were to conduct a one-sided t–test, with an alternate hypothesis that the coefficient is greater than 0, the Pr(t > 3.9550) would be approximately 0, so we can conclude at a 5% level that there is evidence of a positive relationship between Mwage and Unemployment. The coefficient in this regression suggests that for a $1 increase in minimum wage, unemployment would increase by 0.22%, all else held equal.

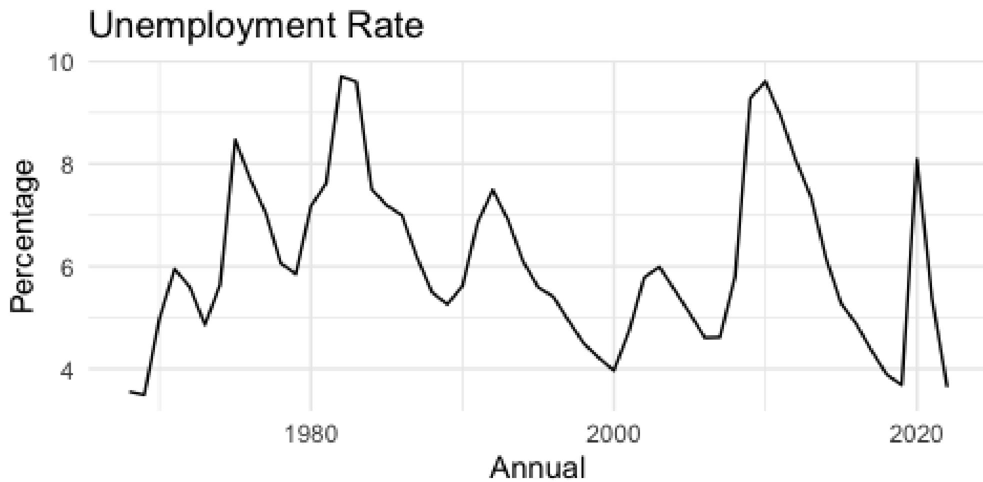
Time series model

Finally, we decided to fit the model using time series analysis for a different approach. We used slightly different data for the annual unemployment rate of the entire United States from 1968-2022 so that we can fit an AR and later ADL model using the Federal Minimum Wage, which is only available as annual data from 1968-2022. In this approach it was not practical to segment the data by state. We start by looking at the time series of Unemployment Rate and making sure the data is stationary:

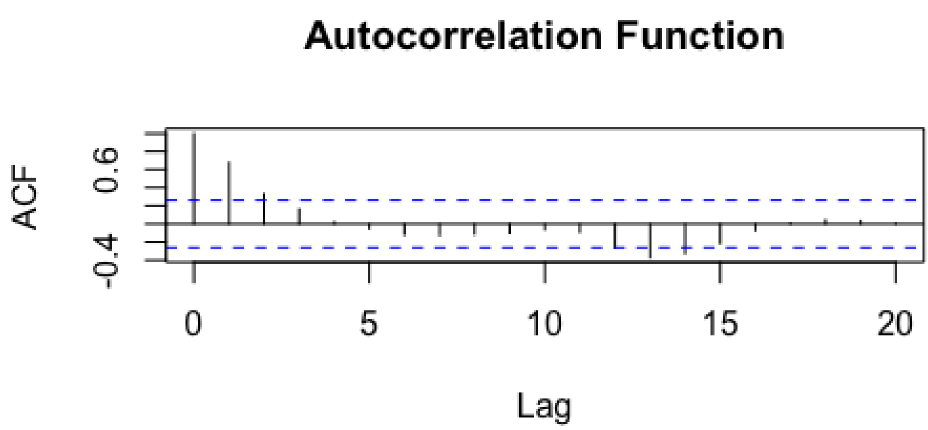
```r
# Read the Excel file
tsdata <- read_excel(path = "UNRate&FedMinWage.xls")

# Print the structure of the imported data (optional)
str(tsdata)
```

```
tibble [55 × 3] (S3: tbl_df/tbl/data.frame)
 $ date      : POSIXct[1:55], format: "1968-01-01" "1969-01-01" ...
 $ FedMinWage: num [1:55] 1.6 1.6 1.6 1.6 1.6 1.6 1.6 1.6 2.3 2.3 ...
 $ UNRATE    : num [1:55] 3.56 3.49 4.98 5.95 5.6 ...
```

```r
# Plot the data
tsdata$date <- as.Date(tsdata$date)
subset_data <- tsdata %>%
  filter(date < as.Date("2022-12-31"))
ggplot(subset_data, aes(x = date, y = UNRATE)) +
  geom_line() +
  labs(title = 'Unemployment Rate',
       x = 'Annual',
       y = 'Percentage') +
  theme_minimal()
```




```
# Plot the autocorrelation function
acf(tsdata$UNRATE, lag.max = 20, na.action = na.omit, main = "Autocorrelation
Function")
```




```
library(tseries)
adf_test <- adf.test(subset_data$UNRATE, alternative = "stationary", k = 4)

# Print the ADF test results
print(adf_test)


    Augmented Dickey-Fuller Test

data:  subset_data$UNRATE
Dickey-Fuller = -2.7713, Lag order = 4, p-value = 0.2637
alternative hypothesis: stationary

# Extract and print the test statistic and p-value
adf_statistic <- adf_test$statistic
```

```
p_value <- adf_test$p.value
cat("ADF Statistic:", adf_statistic, "\n")
```

```
ADF Statistic: -2.771315
```

```
cat("p-value:", p_value, "\n")
```

```
p-value: 0.2637255
```

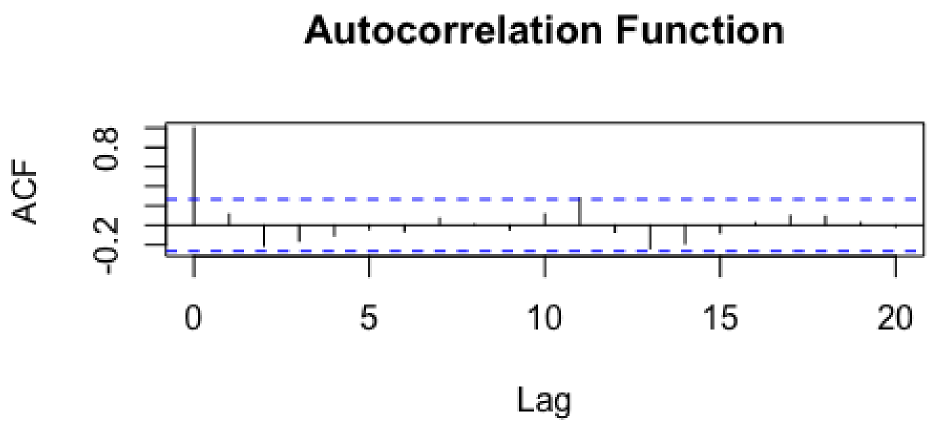
```
# Interpret the results
if (p_value <= 0.05) {
  cat("Reject the null hypothesis. The time series is likely stationary.\n")
} else {
  cat("Fail to reject the null hypothesis. The time series may be non-
stationary.\n")
}
```

```
Fail to reject the null hypothesis. The time series may be non-stationary.
```

To stationarize the data, we take the first difference of the data:

```
# Take difference of the series
tsdata <- tsdata %>%
  mutate(UNRATE_diff = c(NA, diff(UNRATE)))

subset_data <- tsdata %>%
  filter(date < as.Date("2022-12-31"))
ggplot(subset_data, aes(x = date, y = UNRATE_diff)) +
  geom_line() +
  labs(title = 'Difference in Unemployment Rate',
       x = 'Annual',
       y = 'Percentage') +
  theme_minimal()
```

```r
# Plot the autocorrelation function
acf(tsdata$UNRATE_diff, lag.max = 20, na.action = na.omit, main =
"Autocorrelation Function")
```




```r
subset_data <- subset_data[complete.cases(subset_data$UNRATE_diff), ]
adf_test <- adf.test(subset_data$UNRATE_diff, alternative = "stationary", k =
4)

# Print the ADF test results
print(adf_test)


    Augmented Dickey-Fuller Test

data:  subset_data$UNRATE_diff
Dickey-Fuller = -3.9876, Lag order = 4, p-value = 0.01672
alternative hypothesis: stationary

# Extract and print the test statistic and p-value
adf_statistic <- adf_test$statistic
p_value <- adf_test$p.value
cat("ADF Statistic:", adf_statistic, "\n")

ADF Statistic: -3.98759

cat("p-value:", p_value, "\n")

p-value: 0.0167237

# Interpret the results
if (p_value <= 0.05) {
  cat("Reject the null hypothesis. The time series is likely stationary.\n")
} else {
  cat("Fail to reject the null hypothesis. The time series may be non-
stationary.\n")
}
```

```
Reject the null hypothesis. The time series is likely stationary.
```

The differenced data is stationary, so we fit AR(1) to AR(4) models to evaluate which model has the best goodness of fit:

```
library(forecast)

# Fit AR(1) model
ar1_model <- Arima(tsdata$UNRATE_diff, order = c(1, 0, 0), include.constant =
TRUE)
print(summary(ar1_model))
```

```
Series: tsdata$UNRATE_diff
ARIMA(1,0,0) with non-zero mean

Coefficients:
         ar1     mean
      0.1147  -0.0027
s.e.  0.1364   0.1866

sigma^2 = 1.536:  log likelihood = -87.19
AIC=180.39   AICc=180.87   BIC=186.36

Training set error measures:
                      ME     RMSE       MAE      MPE    MAPE      MASE
ACF1
Training set 0.000143696 1.216096 0.8360121 104.8395 105.663 0.8216716
0.027235
```

```
# Fit AR(2) model
ar2_model <- Arima(tsdata$UNRATE_diff, order = c(2, 0, 0), include.constant =
TRUE)
print(summary(ar2_model))
```

```
Series: tsdata$UNRATE_diff
ARIMA(2,0,0) with non-zero mean

Coefficients:
         ar1      ar2    mean
      0.1294  -0.2468  0.0102
s.e.  0.1333   0.1400  0.1449

sigma^2 = 1.478:  log likelihood = -85.69
AIC=179.38   AICc=180.2   BIC=187.34

Training set error measures:
                      ME     RMSE       MAE      MPE     MAPE      MASE
Training set 0.005765762 1.181387 0.8091848 85.80205 111.6317 0.7953045
                    ACF1
Training set -0.02357644
```

```r
# Fit AR(3) model
ar3_model <- Arima(tsdata$UNRATE_diff, order = c(3, 0, 0), include.constant =
TRUE)
print(summary(ar3_model))
```

```
Series: tsdata$UNRATE_diff
ARIMA(3,0,0) with non-zero mean

Coefficients:
         ar1      ar2      ar3    mean
      0.1064  -0.2093  -0.1627  0.0025
s.e.  0.1342   0.1442   0.1696  0.1278

sigma^2 = 1.48:  log likelihood = -85.24
AIC=180.48   AICc=181.73   BIC=190.42

Training set error measures:
                     ME     RMSE       MAE      MPE     MAPE      MASE
Training set 0.01052423 1.170742 0.7960087 66.06932 121.0559 0.7823544
                    ACF1
Training set -0.01340476
```

```r
# Fit AR(4) model
ar4_model <- Arima(tsdata$UNRATE_diff, order = c(4, 0, 0), include.constant =
TRUE)
print(summary(ar4_model))
```

```
Series: tsdata$UNRATE_diff
ARIMA(4,0,0) with non-zero mean

Coefficients:
         ar1      ar2      ar3      ar4    mean
      0.0887  -0.2426  -0.1077  -0.1499  0.0023
s.e.  0.1347   0.1492   0.1803   0.1755  0.1144

sigma^2 = 1.489:  log likelihood = -84.88
AIC=181.75   AICc=183.54   BIC=193.69

Training set error measures:
                     ME   RMSE       MAE      MPE    MAPE      MASE
Training set 0.01302194 1.1624 0.7814111 70.83586 111.243 0.7680072
                     ACF1
Training set -0.007417938
```

```r
bic_ar1 <- AIC(ar1_model)
print(paste('The BIC of the AR(1) model is:', bic_ar1))
```

```
[1] "The BIC of the AR(1) model is: 180.388360910844"
```

```r
bic_ar2 <- AIC(ar2_model)
print(paste('The BIC of the AR(2) model is:', bic_ar2))
```

```
[1] "The BIC of the AR(2) model is: 179.384318989363"
```

```r
bic_ar3 <- AIC(ar3_model)
print(paste('The BIC of the AR(3) model is:', bic_ar3))
```

```
[1] "The BIC of the AR(3) model is: 180.475889050701"
```

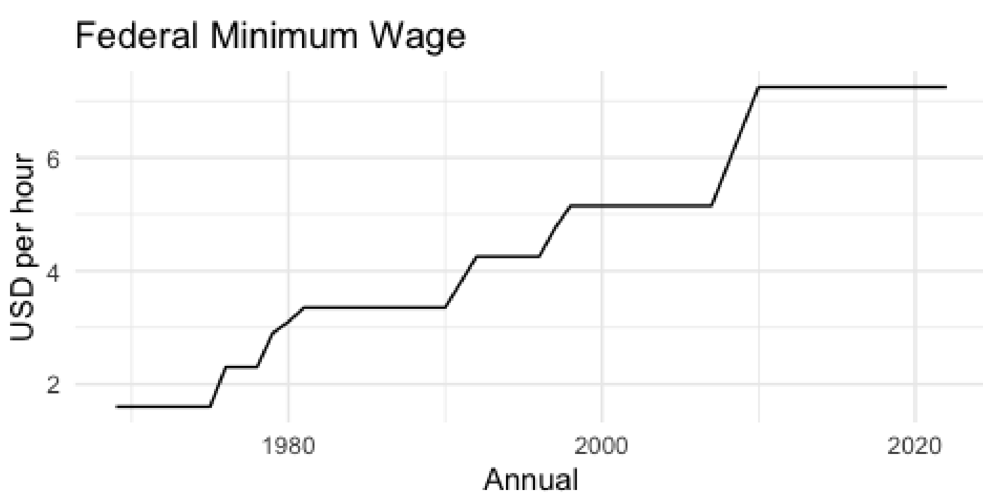
```r
bic_ar4 <- AIC(ar4_model)
print(paste('The BIC of the AR(4) model is:', bic_ar4))
```

```
[1] "The BIC of the AR(4) model is: 181.753941380509"
```

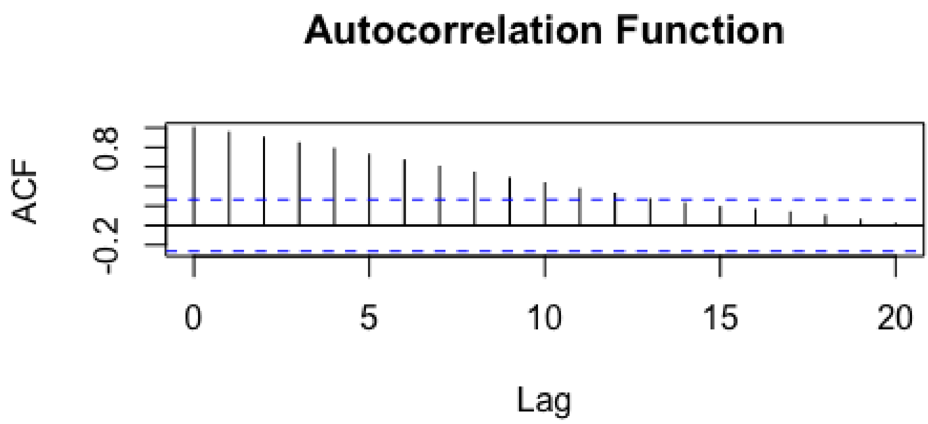
Therefore we select the AR(2) model as the model with the best goodness of fit.

We can now include an ADL model with the Federal Minimum Wage as an explanatory variable.

```r
# Plot the Min Wage data
ggplot(subset_data, aes(x = date, y = FedMinWage)) +
  geom_line() +
  labs(title = 'Federal Minimum Wage',
       x = 'Annual',
       y = 'USD per hour') +
  theme_minimal()
```



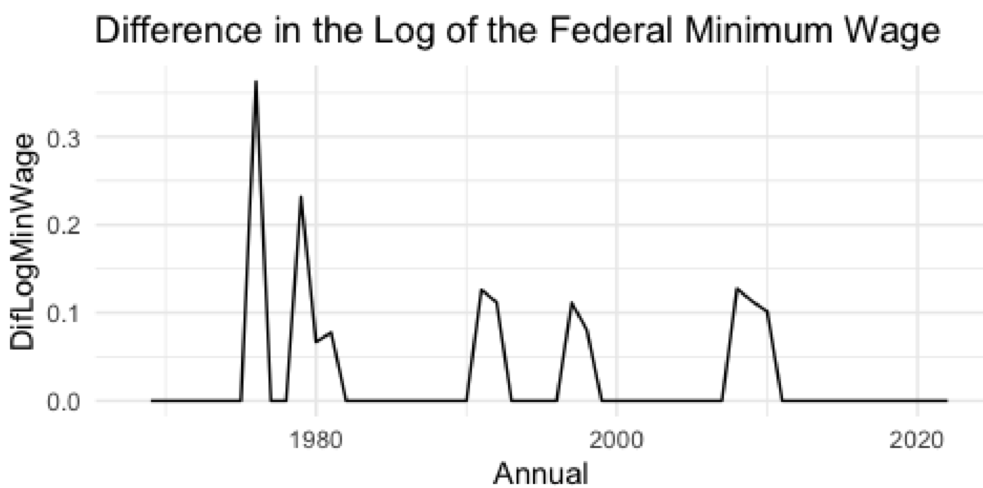

```r
acf(tsdata$FedMinWage, lag.max = 20, na.action = na.omit, main =
"Autocorrelation Function")
```



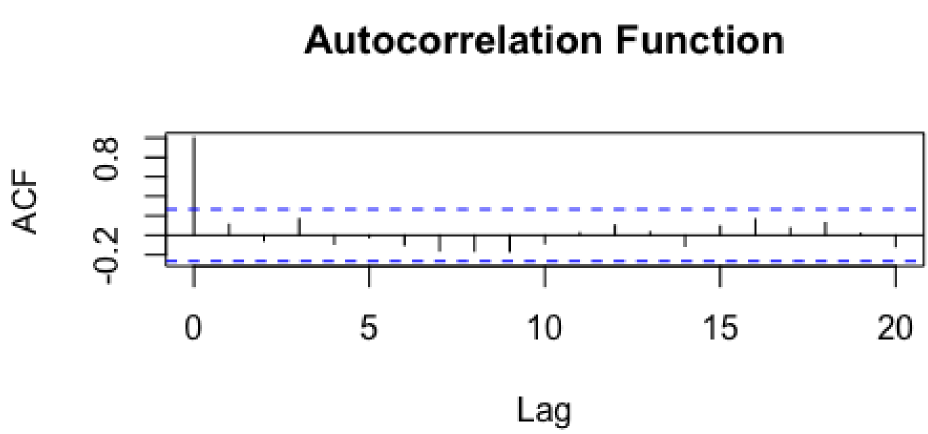

```
adf_test <- adf.test(subset_data$FedMinWage, alternative = "stationary", k =
4)

# Print the ADF test results
print(adf_test)


    Augmented Dickey-Fuller Test

data:  subset_data$FedMinWage
Dickey-Fuller = -2.911, Lag order = 4, p-value = 0.2076
alternative hypothesis: stationary

# Extract and print the test statistic and p-value
adf_statistic <- adf_test$statistic
p_value <- adf_test$p.value
cat("ADF Statistic:", adf_statistic, "\n")

ADF Statistic: -2.91102

cat("p-value:", p_value, "\n")

p-value: 0.2076035

# Interpret the results
if (p_value <= 0.05) {
  cat("Reject the null hypothesis. The time series is likely stationary.\n")
} else {
  cat("Fail to reject the null hypothesis. The time series may be non-
stationary.\n")
}

Fail to reject the null hypothesis. The time series may be non-stationary.
```

First we stationarize the Federal Minimum Wage data by taking the first difference of the log of the raw data, as the original series is not stationary:

```
tsdata <- tsdata %>%
  mutate(DifLogWage = c(NA, diff(log(FedMinWage))))
subset_data <- tsdata %>%
  filter(date < as.Date("2022-12-31"))

ggplot(subset_data, aes(x = date, y = DifLogWage)) +
  geom_line() +
  labs(title = 'Difference in the Log of the Federal Minimum Wage',
       x = 'Annual',
       y = 'DifLogMinWage') +
  theme_minimal()
```




```
acf(tsdata$DifLogWage, lag.max = 20, na.action = na.omit, main =
"Autocorrelation Function")
```




```
# Remove NAs from 'DifLogGDP' column
dif_log_wage_no_na <- na.omit(subset_data$DifLogWage)

# Run ADF test on the cleaned 'DifLogGDP' column
```

```r
adf_test <- adf.test(dif_log_wage_no_na, alternative = "stationary", k = 4)

# Print the ADF test results
print(adf_test)


    Augmented Dickey-Fuller Test

data:  dif_log_wage_no_na
Dickey-Fuller = -3.3974, Lag order = 4, p-value = 0.06563
alternative hypothesis: stationary

# Extract and print the test statistic and p-value
adf_statistic <- adf_test$statistic
p_value <- adf_test$p.value
cat("ADF Statistic:", adf_statistic, "\n")

ADF Statistic: -3.39736

cat("p-value:", p_value, "\n")

p-value: 0.06562743

# Interpret the results
if (p_value <= 0.05) {
  cat("Reject the null hypothesis. The time series is likely stationary.\n")
} else {
  cat("Fail to reject the null hypothesis. The time series may be non-
stationary.\n")
}

Fail to reject the null hypothesis. The time series may be non-stationary.
```

Now we run ADL(2,1) and ADL(2,4). The ADL(2,1) model has the lower BIC:

```r
###ADL model
library(dynlm)
library(MASS)

tsdata <- tsdata %>%
  arrange(date) %>%
  mutate(lag_UNRATE1 = lag(UNRATE, 1),
         lag_UNRATE2 = lag(UNRATE, 2),
         lag_Wage1 = lag(DifLogWage, 1),
         lag_Wage4 = lag(DifLogWage, 4))

subset_data <- tsdata %>%
  filter(date < as.Date("2023-12-31"))

# Fit ADL(2,1) model
adl_21_model <- dynlm(UNRATE_diff ~ lag_UNRATE1 + lag_UNRATE2 + lag_Wage1,
```

```r
data = subset_data)

# Print summary for ADL(2,1) model
summary_21 <- summary(adl_21_model)
bic_21 <- BIC(adl_21_model)
cat("ADL(2,1) Model BIC:", bic_21, "\n")

ADL(2,1) Model BIC: 182.6249

print(summary_21)


Time series regression with "numeric" data:
Start = 1, End = 54

Call:
dynlm(formula = UNRATE_diff ~ lag_UNRATE1 + lag_UNRATE2 + lag_Wage1,
    data = subset_data)

Residuals:
    Min      1Q  Median      3Q     Max
-2.4968 -0.7670 -0.0320  0.5111  3.8632

Coefficients: (1 not defined because of singularities)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.7512     0.6296  -2.782  0.00756 **
lag_UNRATE1   0.2838     0.1015   2.797  0.00726 **
lag_UNRATE2       NA         NA      NA       NA
lag_Wage1     0.5183     2.4439   0.212  0.83291
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.165 on 51 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.1444,    Adjusted R-squared:  0.1108
F-statistic: 4.304 on 2 and 51 DF,  p-value: 0.01875

# Fit ADL(2,4) model
adl_24_model <- dynlm(UNRATE ~ lag_UNRATE1 + lag_UNRATE2 + lag_Wage4, data =
subset_data)

# Print summary for ADL(3,4) model
summary_24 <- summary(adl_24_model)
bic_24 <- BIC(adl_24_model)
cat("ADL(2,4) Model BIC:", bic_24, "\n")

ADL(2,4) Model BIC: -3645.672

print(summary_24)
```

```
Time series regression with "numeric" data:
Start = 1, End = 54

Call:
dynlm(formula = UNRATE ~ lag_UNRATE1 + lag_UNRATE2 + lag_Wage4,
    data = subset_data)

Residuals:
                  Min                     1Q                 Median
-0.0000000000000032677  0.000000000000000212  0.000000000000000624
                   3Q                    Max
 0.000000000000000979  0.0000000000000003978

Coefficients: (1 not defined because of singularities)
                          Estimate             Std. Error
(Intercept) -0.00000000000000136002  0.00000000000000025381
lag_UNRATE1  1.00000000000000022204  0.00000000000000004091
lag_UNRATE2                      NA                     NA
lag_Wage4    0.00000000000000056838  0.00000000000000098531
                         t value            Pr(>|t|)
(Intercept)               -5.358          0.00000203 ***
lag_UNRATE1 24444795306401204.000 < 0.0000000000000002 ***
lag_UNRATE2                   NA                  NA
lag_Wage4                  0.577               0.567
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0000000000000004698 on 51 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:      1, Adjusted R-squared:      1
F-statistic: 3.158e+32 on 2 and 51 DF,  p-value: < 0.00000000000000022
```

This model suggests that the federal minimum wage does not have a significant effect on future predictions of the unemployment rate. Ultimately, we felt that the time series analysis would be biased by many factors such as education, types of employers, and other legislation that would vary by state. For that reason, the panel data with fixed effects seems to be the preferable model.

Conclusion

In conclusion, the best fit model for our data set was the panel regression with time effects and income, education, and GDP. This model provided significant evidence that there is a positive relationship between minimum wage and unemployment in a given state at a given time at the 5% confidence level, holding the other factors constant. This conclusion seemingly aligns with the more recent results published by Meer and West, as well as Clemens and Wither. However, it does not necessarily refute Card and Krueger because our analysis was performed on a large time scale, and theirs had a much more narrow scope. We can conclude holistically that raising minimum wage might lead to increases in unemployment, but this relationship can change under specific circumstances of time, state, and macroeconomic conditions.

Works Cited

Card, David, and Alan Krueger. "Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania." *The American Economic Review* 84, no.4 (September 1994): 772–93.

Meer, Jonathan, and Jeremy West. "Effects of the Minimum Wage on Employment Dynamics." National Bureau of Economic Research, Working Paper No. 19262 (August 2013).

Clemens, Jeffrey, and Michael Wither. "The Minimum Wage and the Great Recession: Evidence of Effects on the Employment and Income Trajectories of Low-Skilled Workers." National Bureau of Economic Research, Working Paper No. 20724 (December 2014).