# Forecasting Employment in Daytona-Deltona Florida

*Kahlil Wehmeyer*

*3/18/2019*

## Contents



## Abstract

*The purpose of this project is to forecast the March non seasonally adjusted estimates of average hourly earnings, average weekly hours, average weekly earnings, total employment and total weekly earnings for the Deltona Daytona metropolitan statistical area*

# Introduction

We will be predicting Average Hourly Earnings, Average Weekly Hours and Total Employment.

From there we can calculate average weekly earnings as such:

$$AverageWeeklyEarnings = AverageHourlyEarnings \times AverageWeeklyHours$$

Total weekly earnings will be calculated as such:

$$TotalWeeklyEarnings = TotalEmployment \times AverageWeeklyEarnings$$

## Data

The data we will be using is sourced from multiple CSVs downloaded from the FRED.ORG website.

- All Employees Federal Government
- All Employees Health Care
- All Employees Local Government
- All Employees Retail Trade
- All Employees Total Private
- Average Weekly Earnings All Private
- Average Weekly Hours All Private

# Data

Here we will import all the data and rename the variables to something more useful. Initially columns have names starting with `SMU` that are followed by a 15 digit number.

| D1_Federal_Employees | D1_Healthcare_Employees | D1_Local_Gov_Employees | D1_Retail_Employees | D1_All_E |
|---|---|---|---|---|
| 0.1 | 0.1 | 0.4 | 0.1 | |
| 0.0 | 0.0 | 0.1 | 0.4 | |
| 0.0 | 0.0 | 0.0 | 0.6 | |
| 0.0 | 0.0 | 0.0 | 0.1 | |
| 0.0 | 0.0 | -0.2 | -0.5 | |
| 0.0 | 0.0 | 0.1 | 0.1 | |

## Summary Statistics

Table 2: Table continues below

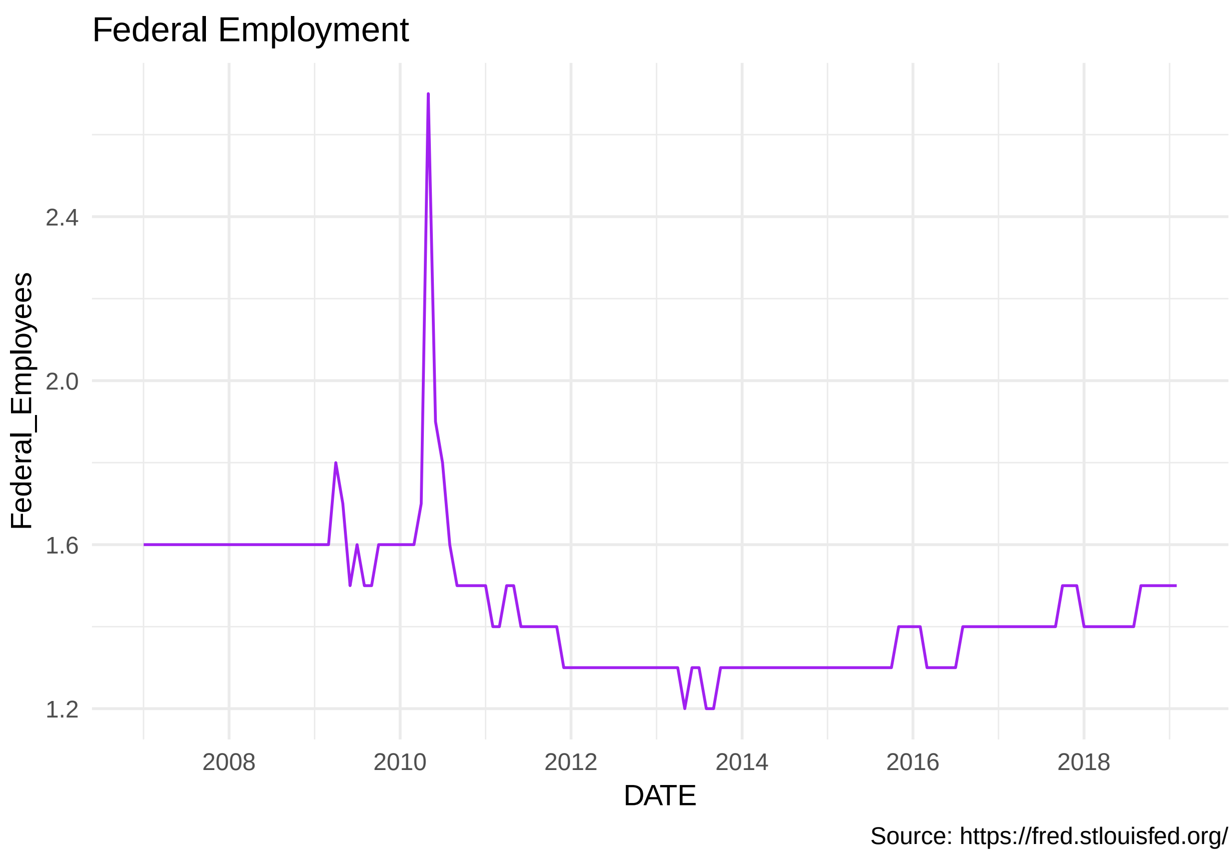
| DATE | Federal_Employees | Healthcare_Employees |
|---|---|---|
| Min. :2007-01-01 | Min. :1.200 | Min. : 8.100 |
| 1st Qu.:2010-01-08 | 1st Qu.:1.300 | 1st Qu.: 8.300 |
| Median :2013-01-16 | Median :1.400 | Median : 8.500 |
| Mean :2013-01-15 | Mean :1.445 | Mean : 8.741 |
| 3rd Qu.:2016-01-24 | 3rd Qu.:1.600 | 3rd Qu.: 8.875 |
| Max. :2019-02-01 | Max. :2.700 | Max. :10.100 |

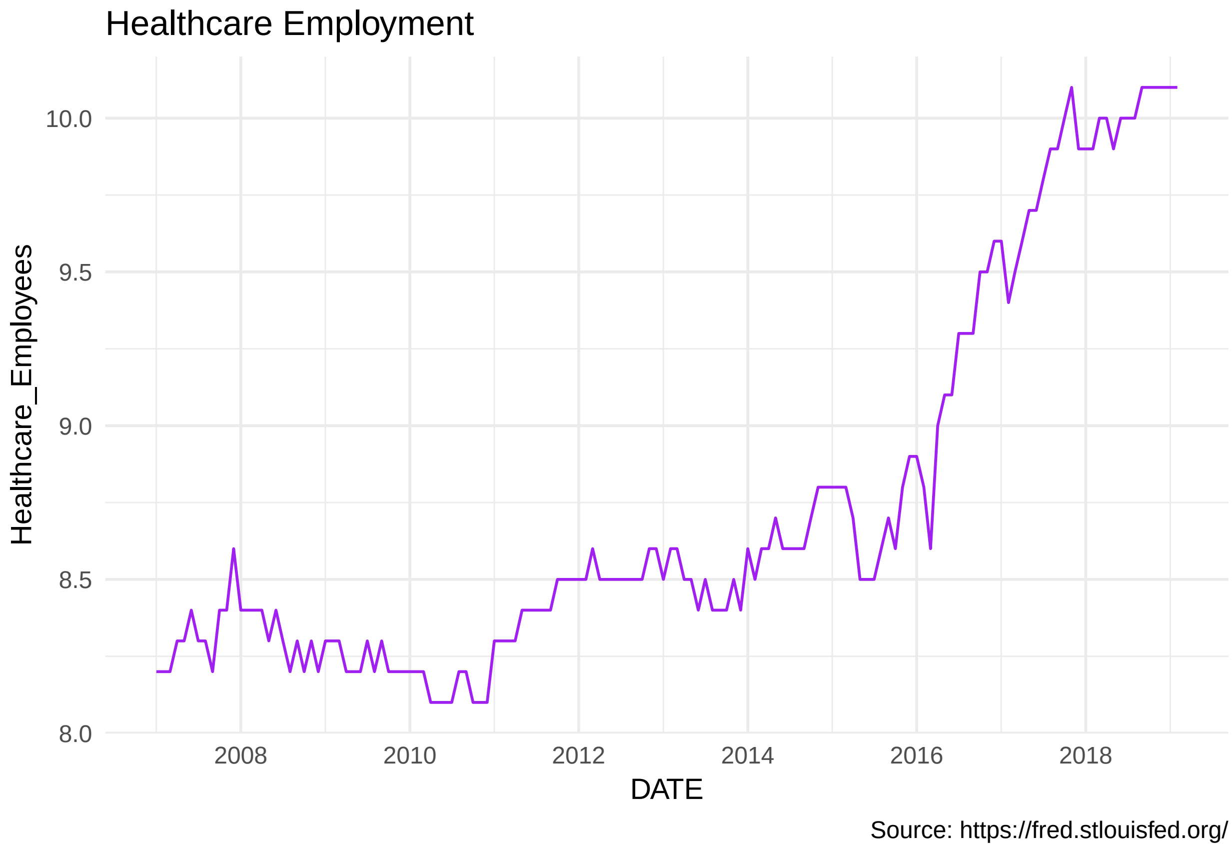
Table 3: Table continues below

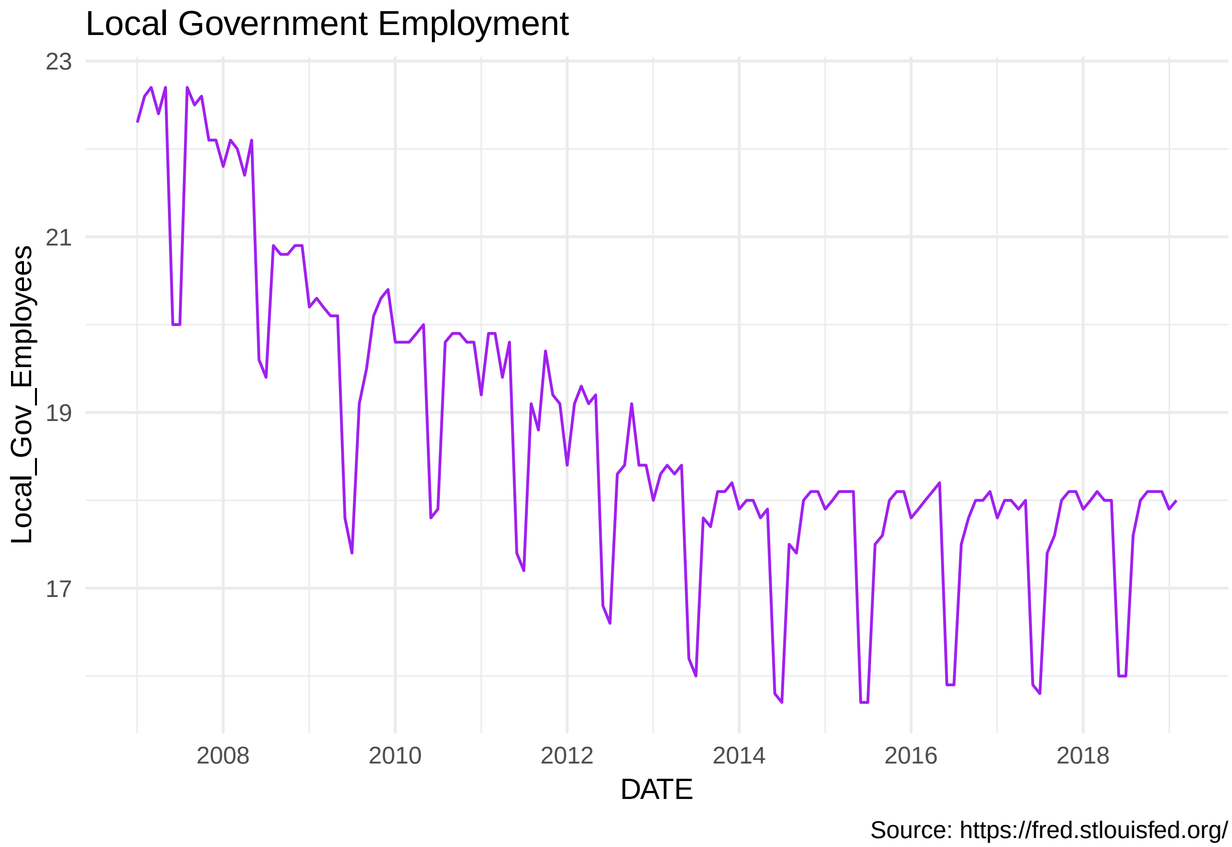
| Retail_Employees | All_Employees | Average_Weekly_Hours |
|---|---|---|
| Min. :25.90 | Min. :144.0 | Min. :30.60 |
| 1st Qu.:26.70 | 1st Qu.:149.7 | 1st Qu.:32.90 |
| Median :27.80 | Median :158.6 | Median :33.80 |
| Mean :28.16 | Mean :160.3 | Mean :33.82 |
| 3rd Qu.:29.40 | 3rd Qu.:170.9 | 3rd Qu.:34.67 |
| Max. :31.90 | Max. :183.3 | Max. :36.50 |

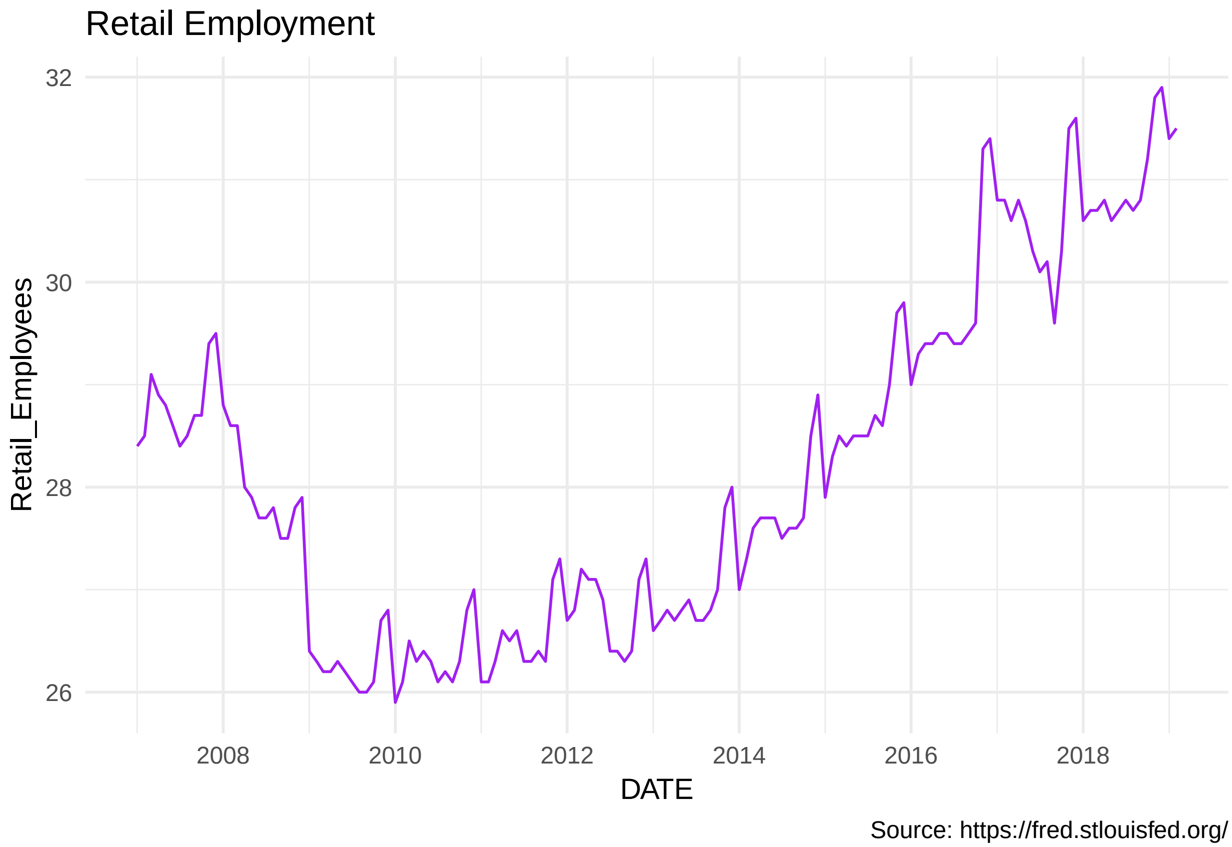
| Average_Hourly_Earnings | Average_Weekly_Earnings | Total_Weekly_Earnings |
|---|---|---|
| Min. :16.46 | Min. :530.9 | Min. : 79215 |
| 1st Qu.:18.08 | 1st Qu.:600.8 | 1st Qu.: 92878 |
| Median :19.05 | Median :646.6 | Median :105354 |
| Mean :19.78 | Mean :670.2 | Mean :107518 |
| 3rd Qu.:21.59 | 3rd Qu.:722.3 | 3rd Qu.:124185 |
| Max. :26.11 | Max. :930.1 | Max. :147511 |

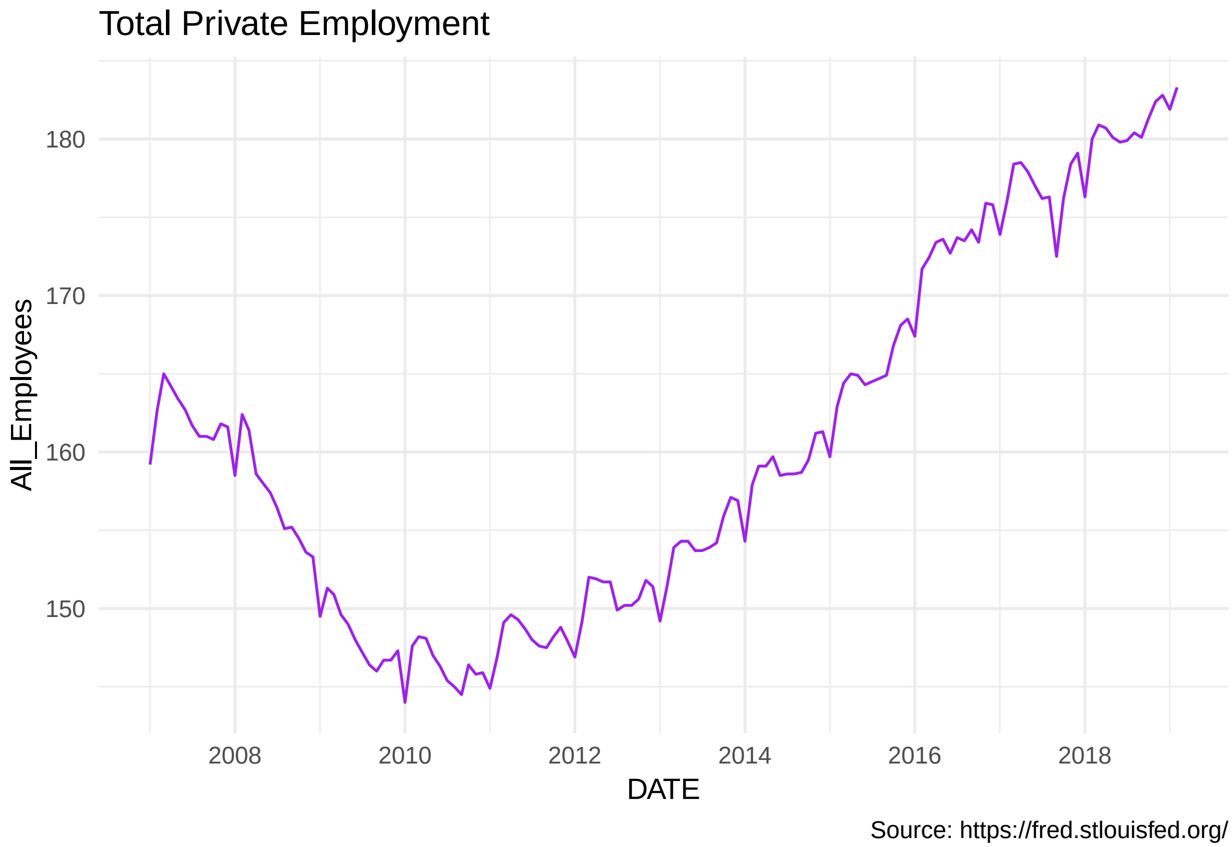
### Time Series Lines

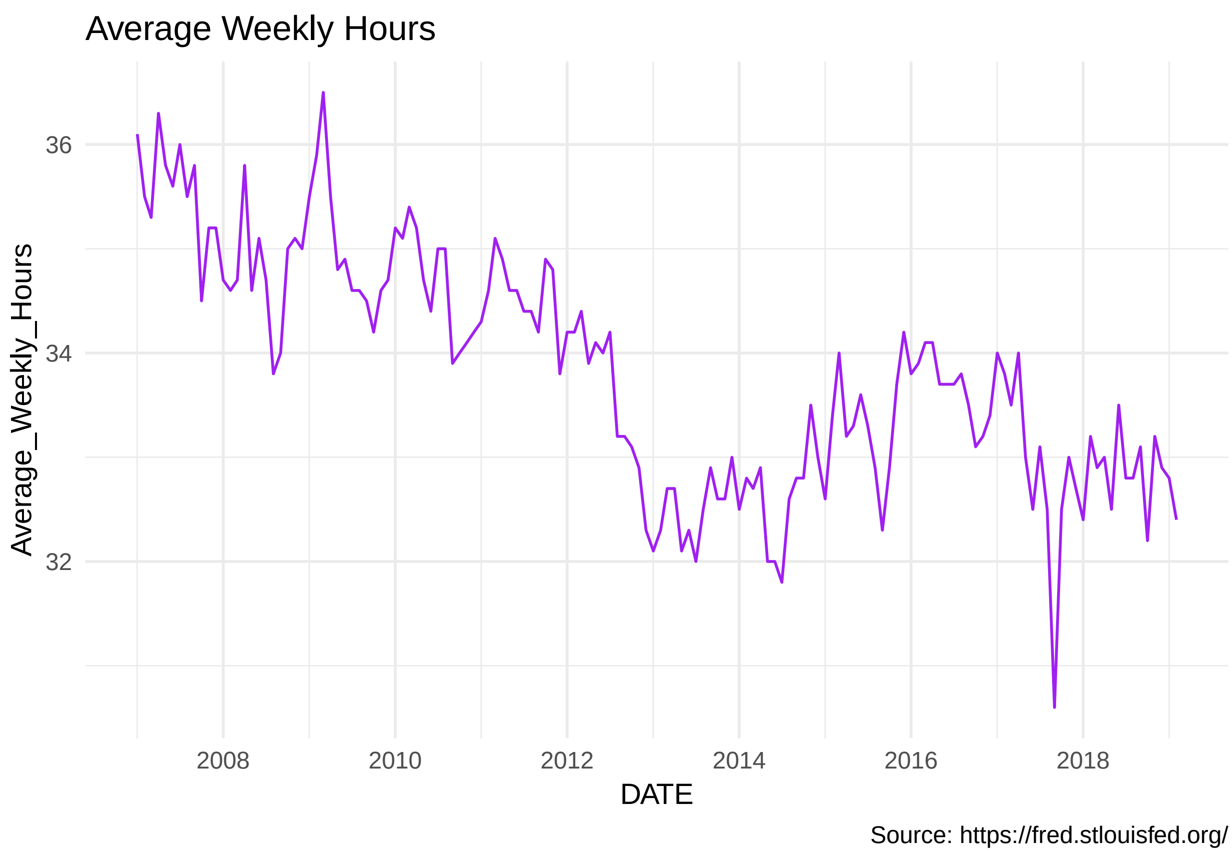



Healthcare Employment
Healthcare_Employees
10.0
9.5
9.0
8.5
8.0
2008
2010
2012
2014
2016
2018
DATE
Source: https://fred.stlouisfed.org/

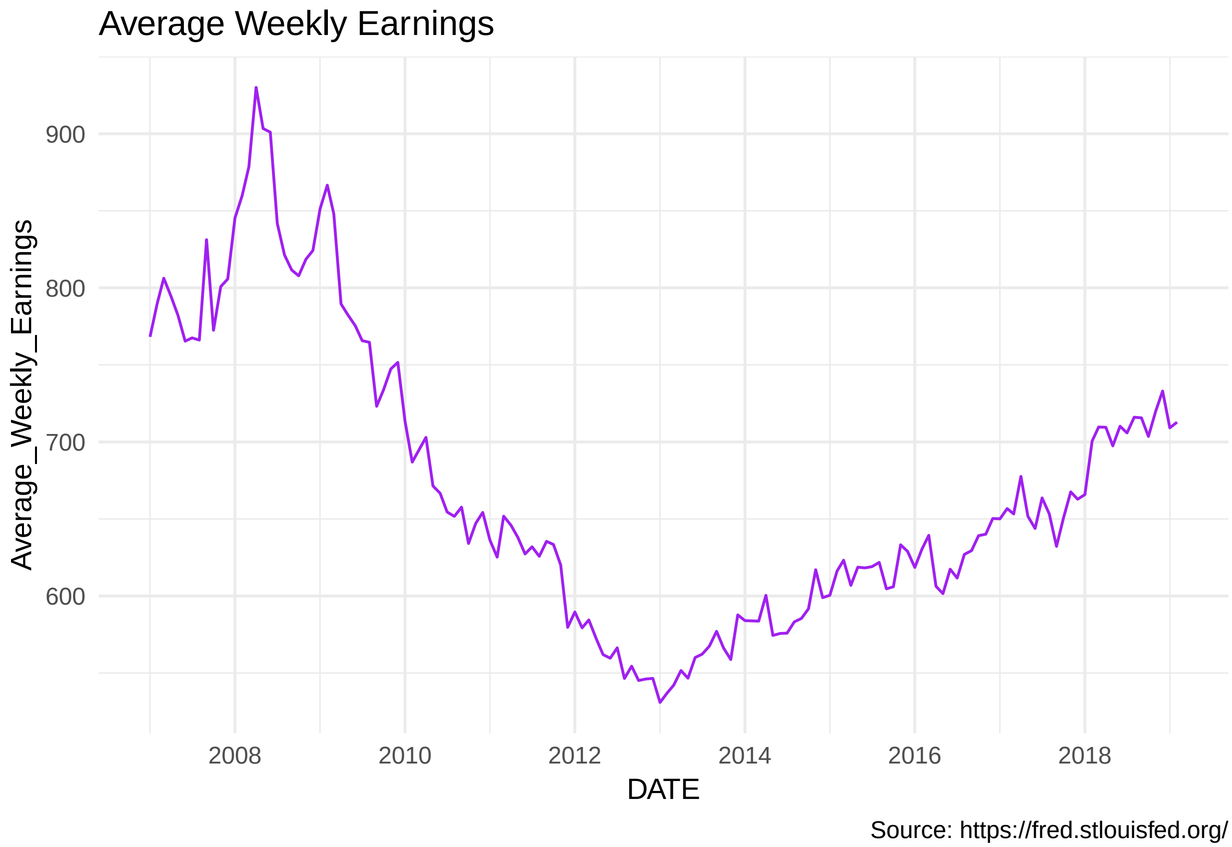

Local Government Employment
23
21
19
17
Local_Gov_Employees
2008
2010
2012
2014
2016
2018
DATE
Source: https://fred.stlouisfed.org/

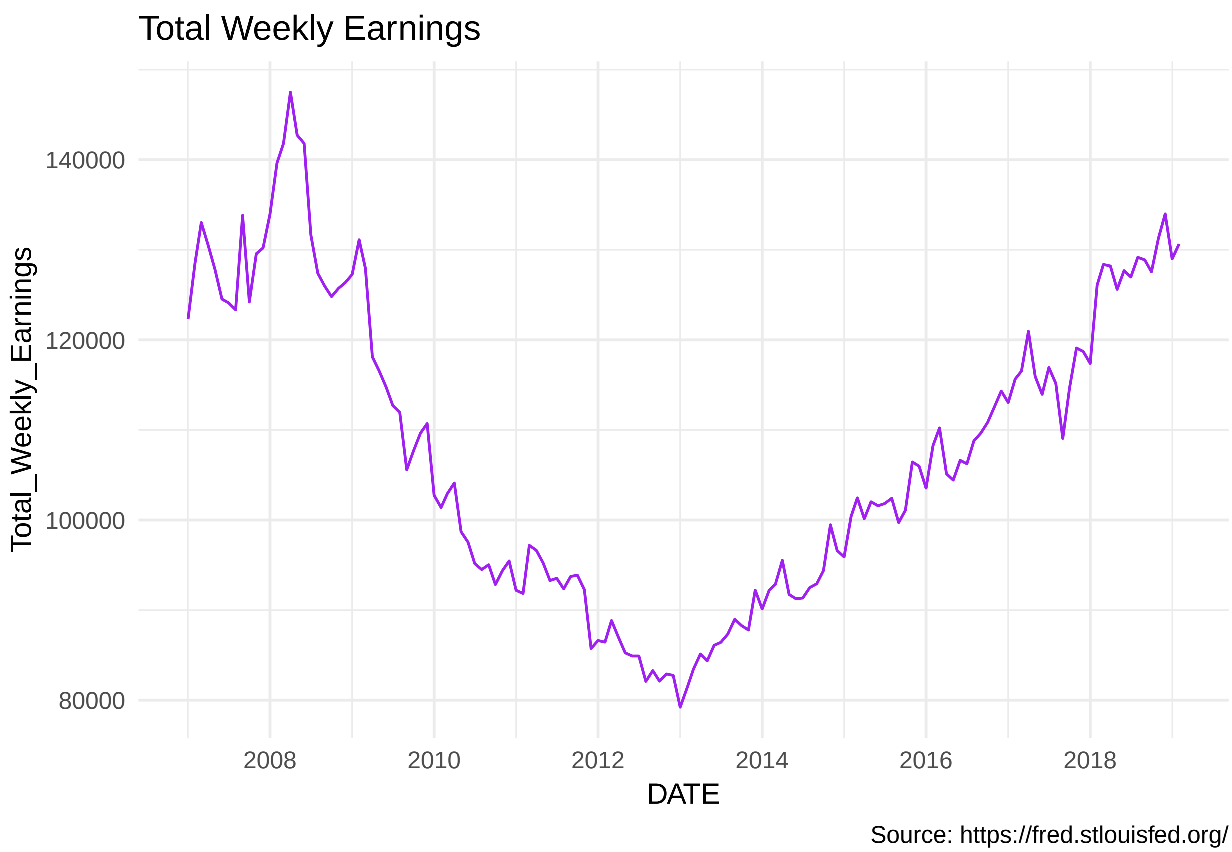

Retail Employment
Retail_Employees
32
30
28
26
2008
2010
2012
2014
2016
2018
DATE
Source: https://fred.stlouisfed.org/


Total Private Employment
All_Employees
180
170
160
150
2008
2010
2012
2014
2016
2018
DATE
Source: https://fred.stlouisfed.org/


Average Weekly Hours
Average_Weekly_Hours
36
34
32
2008
2010
2012
2014
2016
2018
DATE
Source: https://fred.stlouisfed.org/

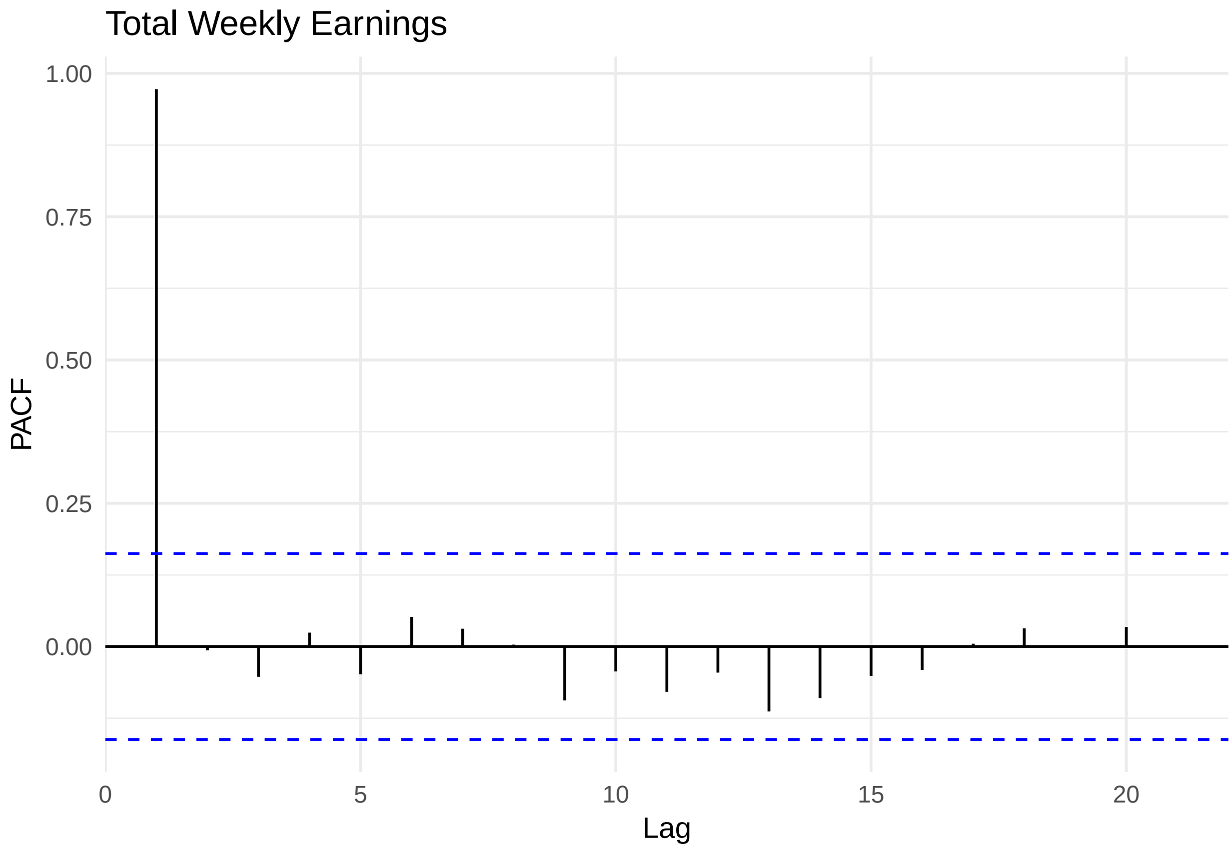

Average Hourly Earnings
26
24
22
20
18
16
Average_Hourly_Earnings
2008
2010
2012
2014
2016
2018
DATE
Source: https://fred.stlouisfed.org/

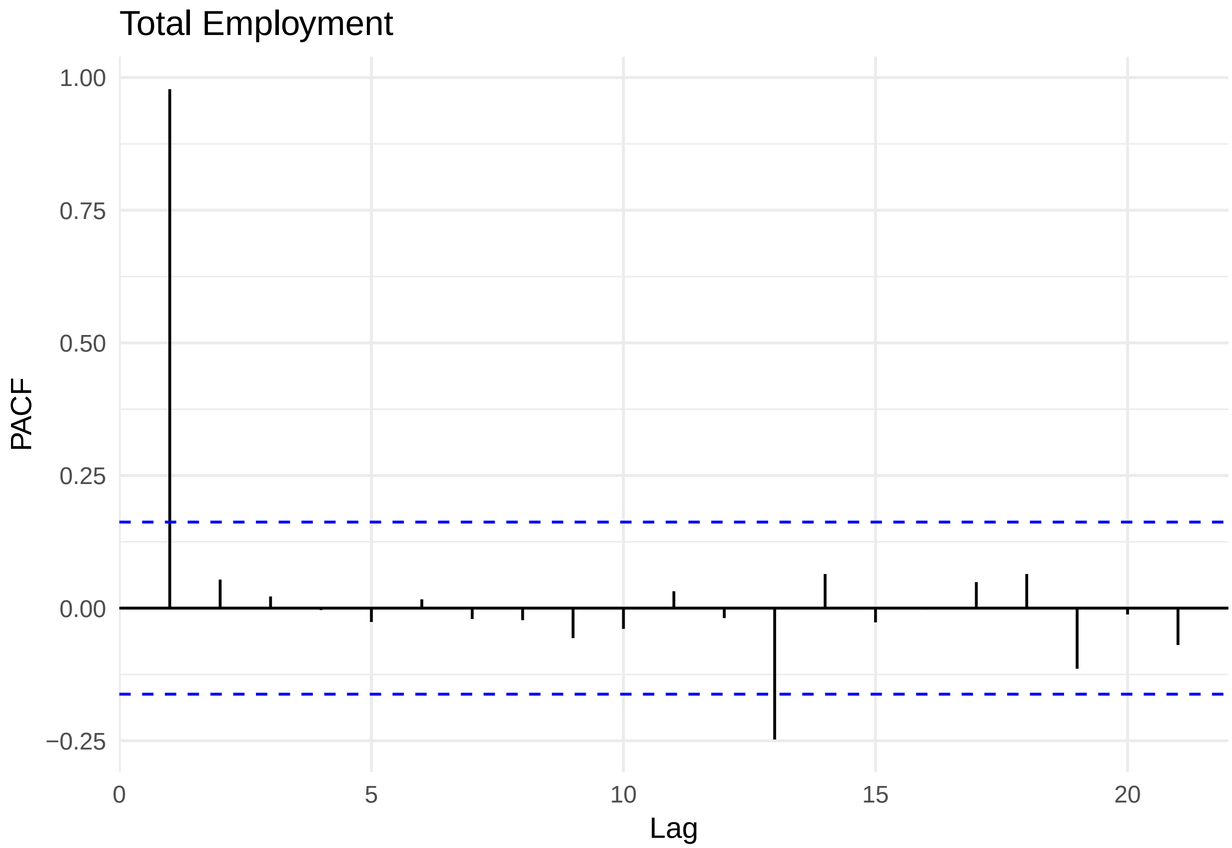

Average Weekly Earnings
Average_Weekly_Earnings
900
800
700
600
2008
2010
2012
2014
2016
2018
DATE
Source: https://fred.stlouisfed.org/

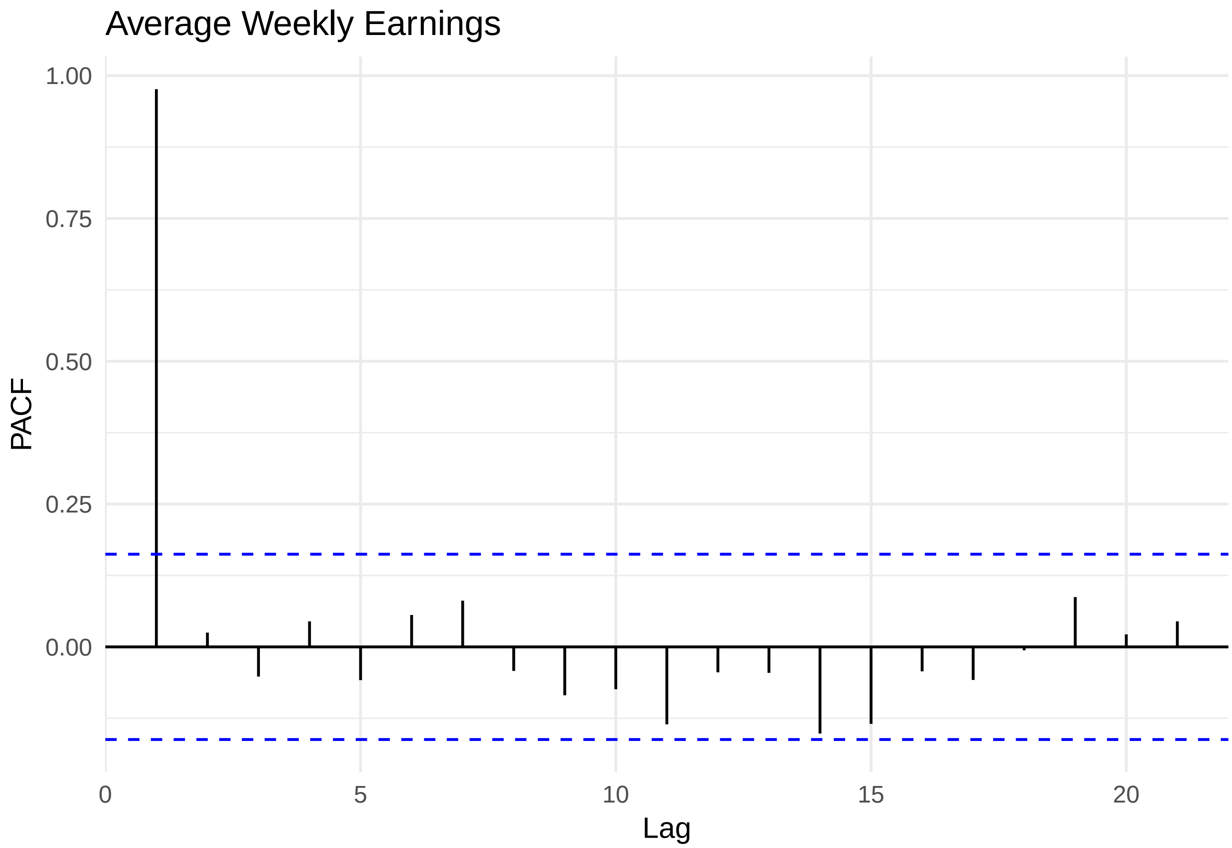



### PACS

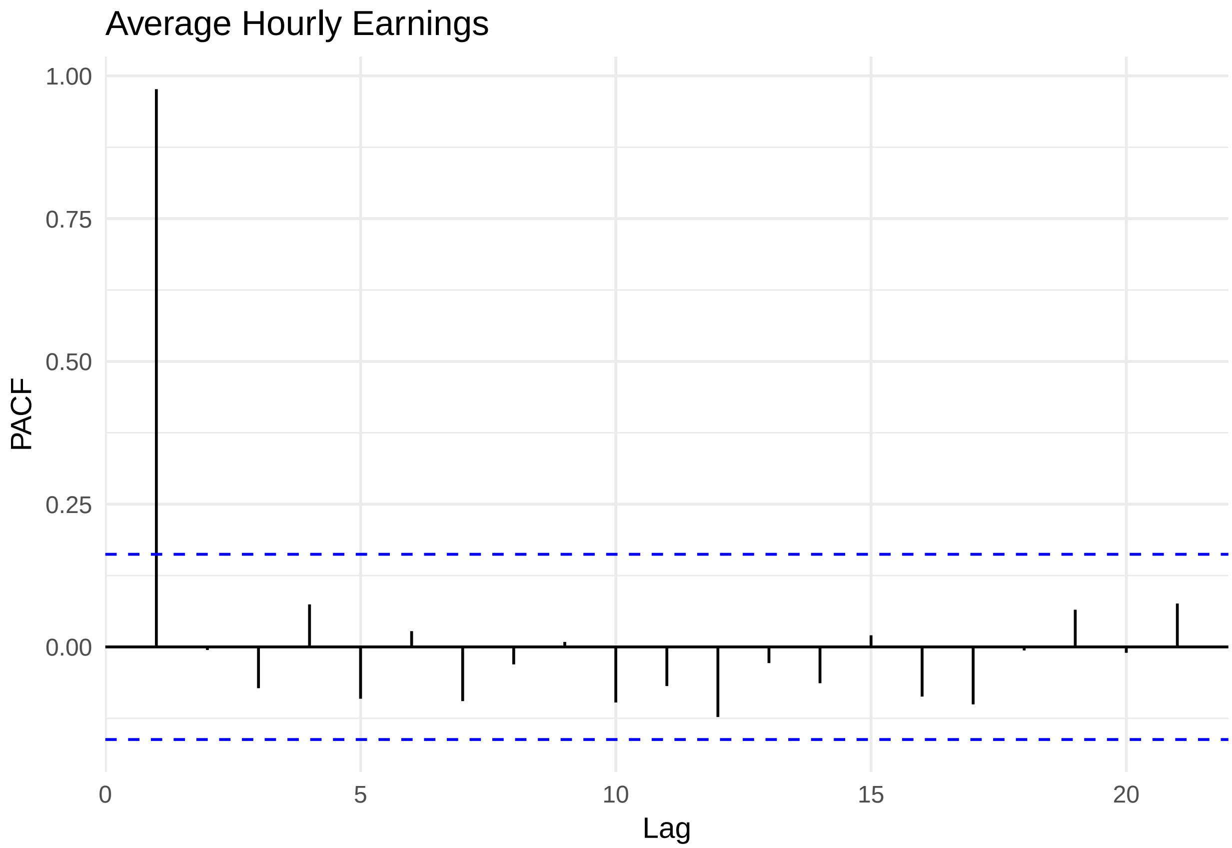
Partial auto correlation is a measure of how stationary a time series value is with it's own lags. When regressing we need to strive for stationarity. Since we don't want things that might have happened in the past to indefinitely effect our forecast of the the future. The common method for addressing stationarity is differencing the data. This mean observing the change in values over time rather than the values themselves.

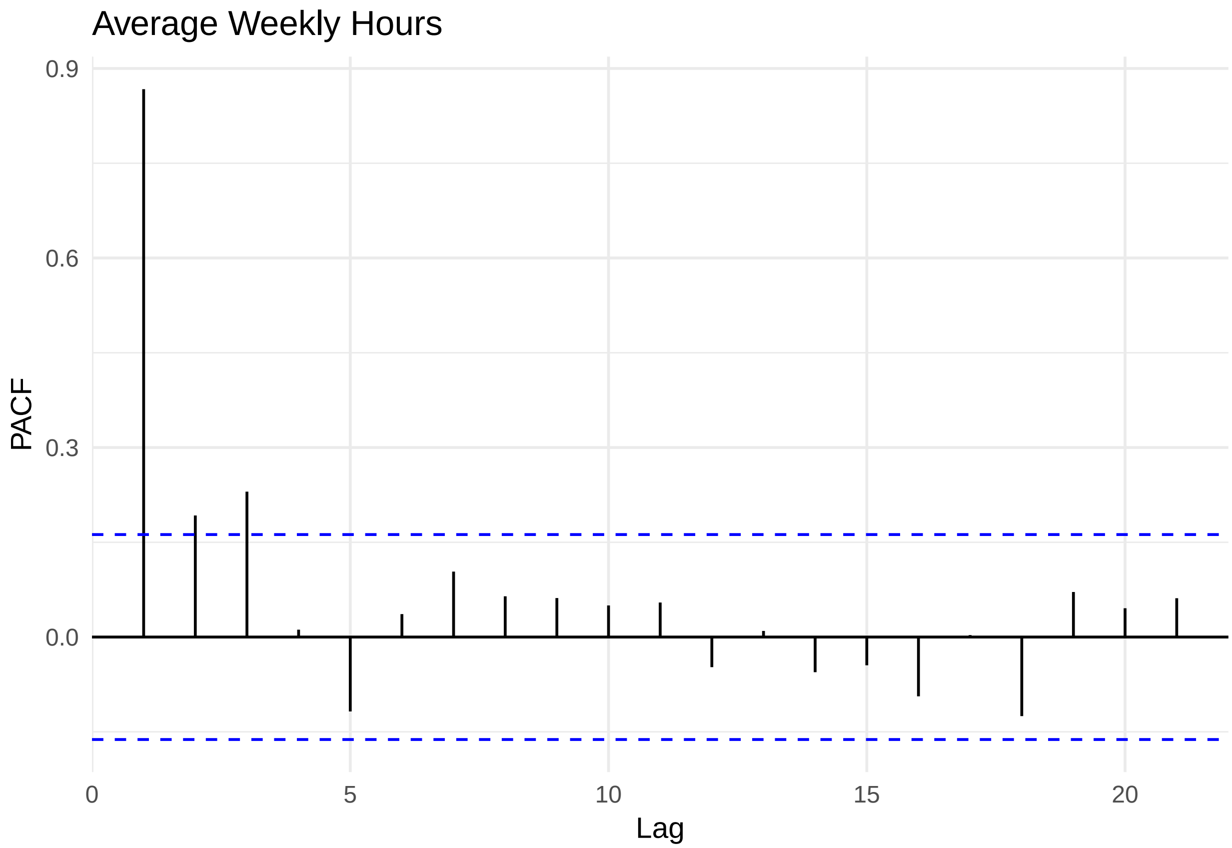

Total Weekly Earnings
1.00
0.75
0.50
0.25
0.00
PACF
0
5
10
15
20
Lag


Total Employment
1.00
0.75
0.50
0.25
0.00
−0.25
PACF
0
5
10
15
20
Lag



Average Weekly Earnings
1.00
0.75
0.50
0.25
0.00
PACF
0
5
10
15
20
Lag


Average Hourly Earnings
PACF
1.00
0.75
0.50
0.25
0.00
0
5
10
15
20
Lag



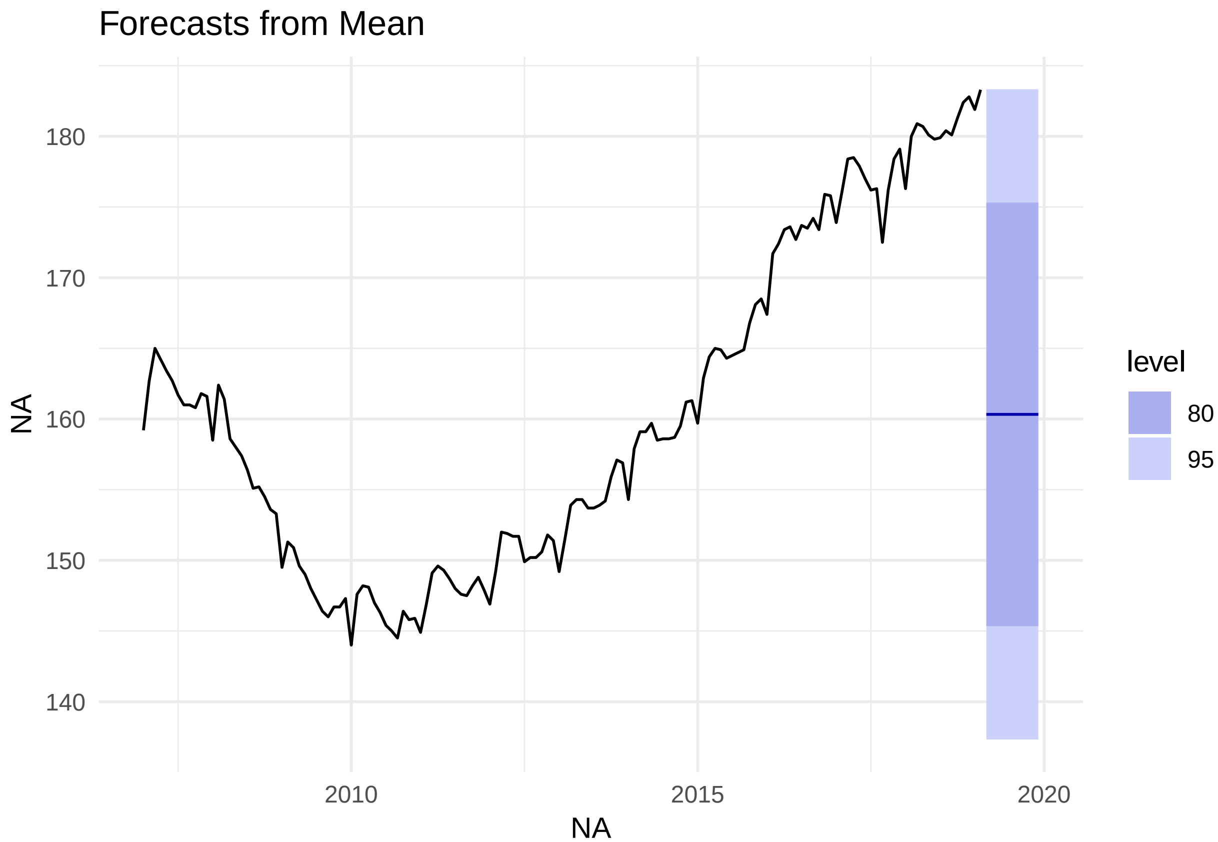

## Preliminary Models

### Simple Models

#### Average Method

Forecasts of all future values are equal to the mean of the historical data. $\hat{y}_{T+h|T} = \bar{y} = (y_1 + \cdots + y_T)/T$

```
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Mar 2019       160.3301 145.3416 175.3187 137.3195 183.3407
## Apr 2019       160.3301 145.3416 175.3187 137.3195 183.3407
## May 2019       160.3301 145.3416 175.3187 137.3195 183.3407
## Jun 2019       160.3301 145.3416 175.3187 137.3195 183.3407
```

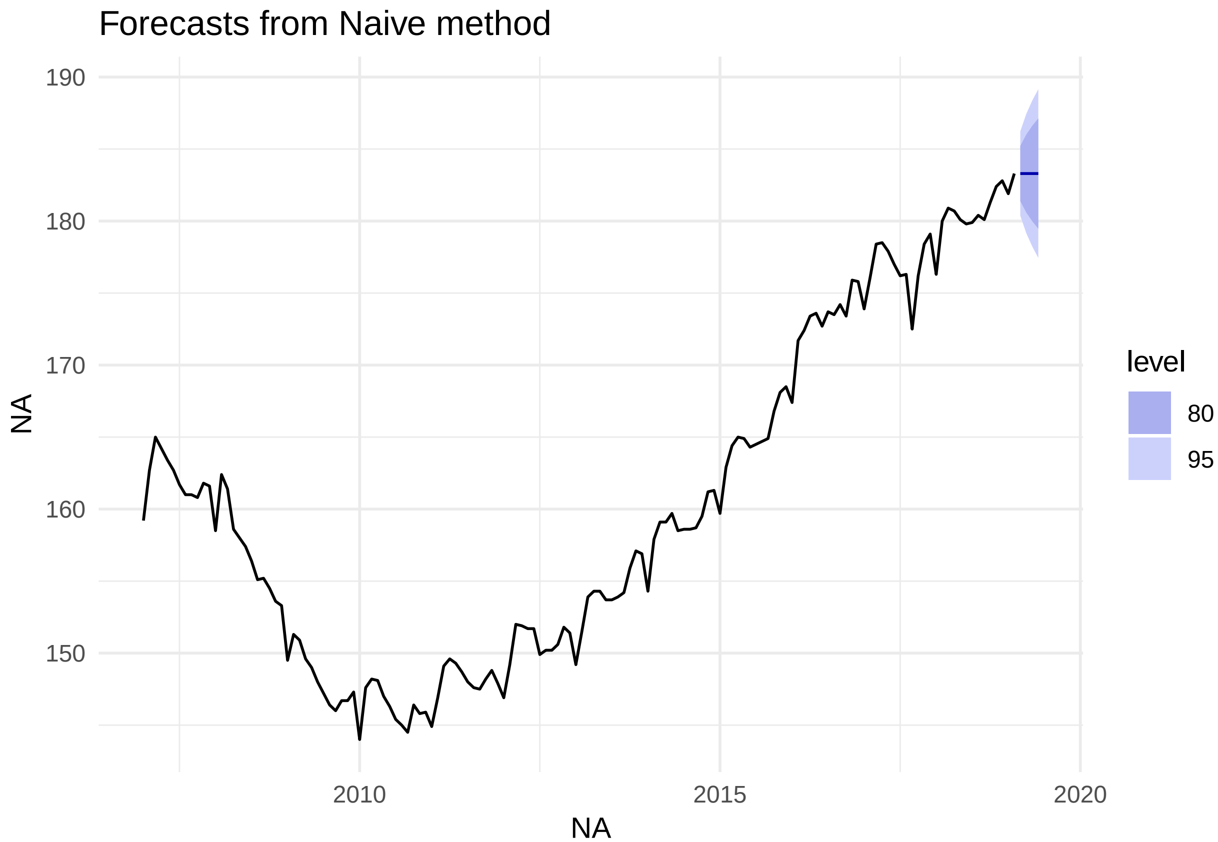



Doesn't look like the ideal method.

**Naive Method**

All forecasts are set to the value of the last observation. $\hat{y}_{T+h|T} = y_T$

```
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Mar 2019          183.3 181.3848 185.2152 180.3710 186.2290
## Apr 2019          183.3 180.5915 186.0085 179.1577 187.4423
## May 2019          183.3 179.9828 186.6172 178.2268 188.3732
## Jun 2019          183.3 179.4696 187.1304 177.4420 189.1580
```

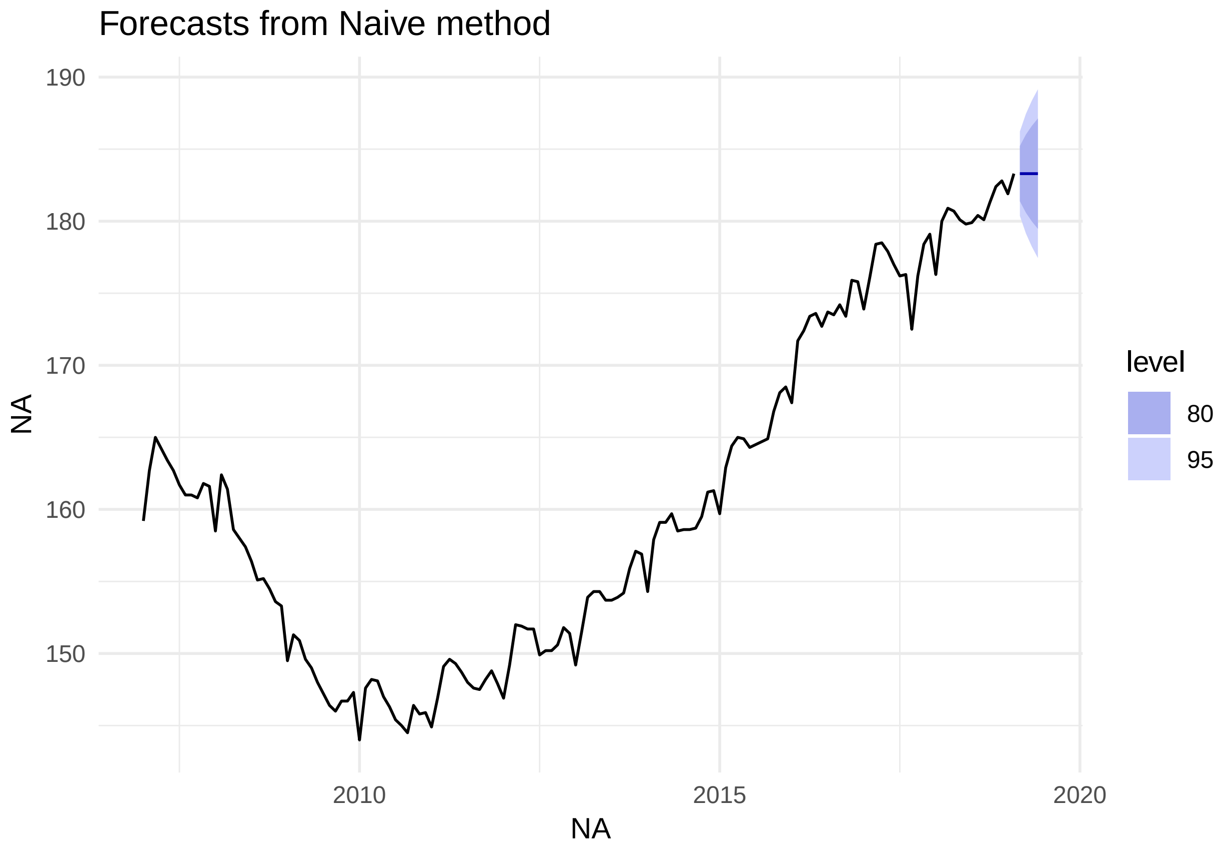



This looks more reasonable but the forecast doesn't change no matter how many period we go out.

**Seasonal Naive Method**

A similar method is useful for highly seasonal data. Here we set each forecast to be the equal to the last observed value from the same season of the year. $\hat{y}_{T+h|T} = y_{T+h-m(k+1)}$,

```
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Mar 2019          180.9 174.1838 187.6162 170.6285 191.1715
## Apr 2019          180.7 173.9838 187.4162 170.4285 190.9715
## May 2019          180.1 173.3838 186.8162 169.8285 190.3715
## Jun 2019          179.8 173.0838 186.5162 169.5285 190.0715
```

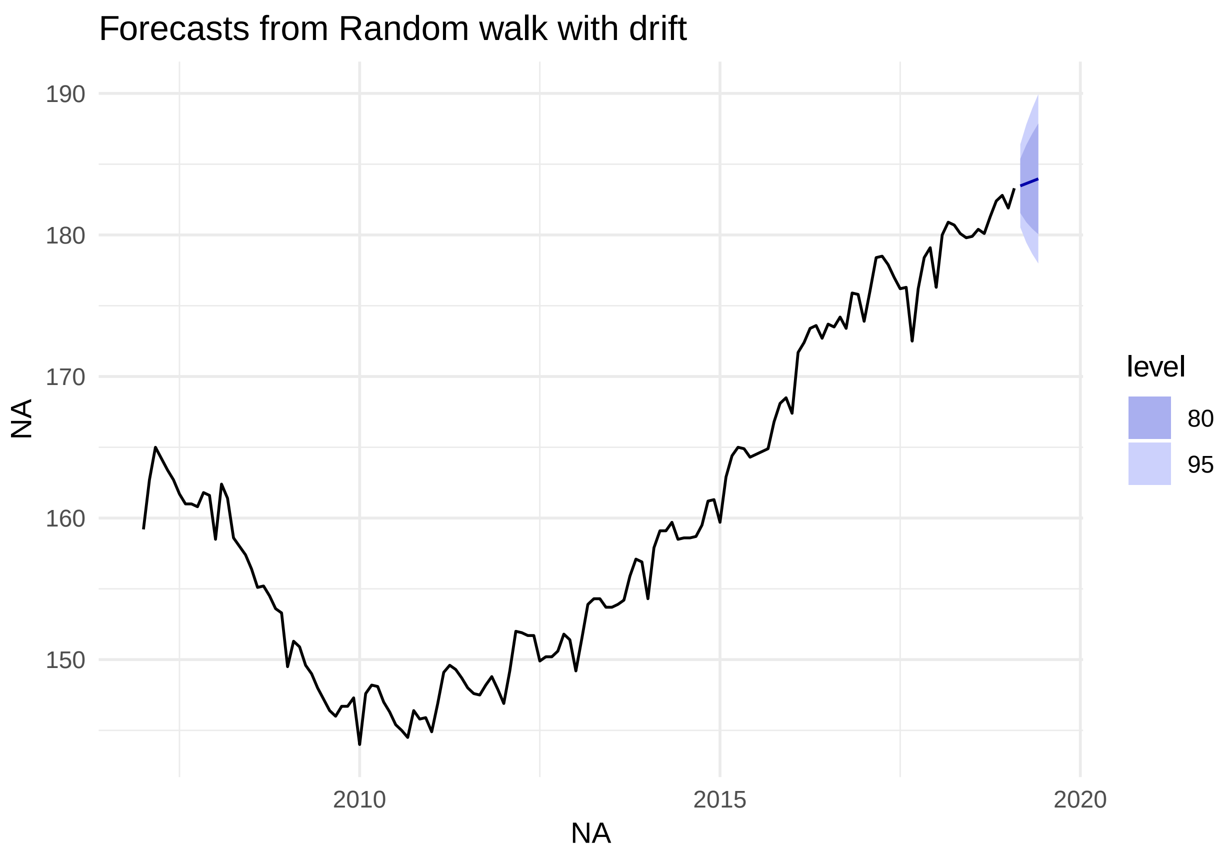



This is a very marginal improvement with the forecasts only differing by decimals.

**Drift Method**

This is a variation on the naive method which allows the forecast to increase or decrease over time. The amount of change, drift, is set to be the average change seen in the historical data. $\hat{y}_{T+h|T} = y_T + \frac{h}{T-1}\sum_{t=2}^{T}(y_t - y_{t-1}) = y_T + h\left(\frac{y_T - y_1}{T-1}\right)$

```
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Mar 2019       183.4662 181.5497 185.3827 180.5352 186.3972
## Apr 2019       183.6324 180.9036 186.3612 179.4591 187.8058
## May 2019       183.7986 180.4340 187.1632 178.6529 188.9443
## Jun 2019       183.9648 180.0539 187.8757 177.9836 189.9461
```




Another marginal improvement.

## Linear Models

Here we will start exploring some models that incorporate both purely auto regressive techniques as well as dynamic models. Something that was not shown in the data until this point is that we have created lagged and differenced values for all the variables.

### Average Weekly Hours

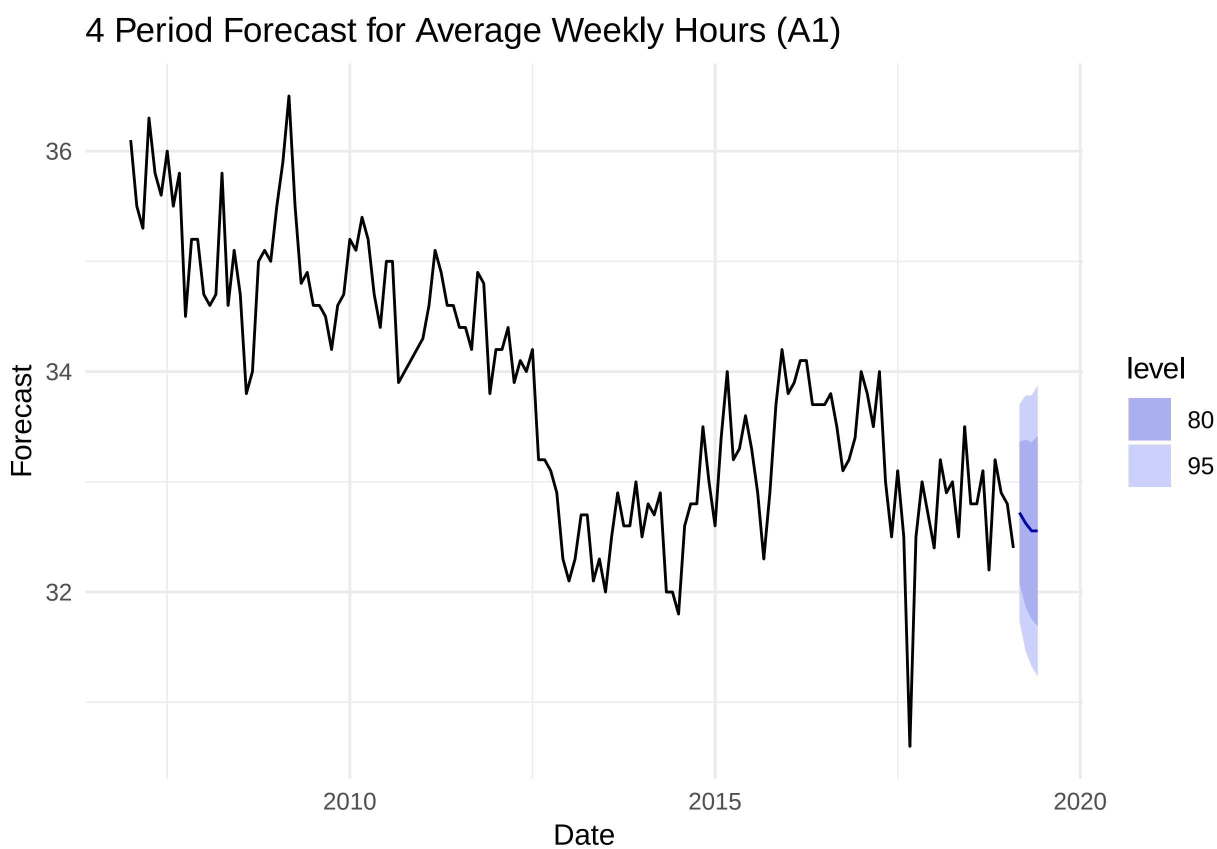
#### ARIMA of Average Weekly Hours (12,1,0)

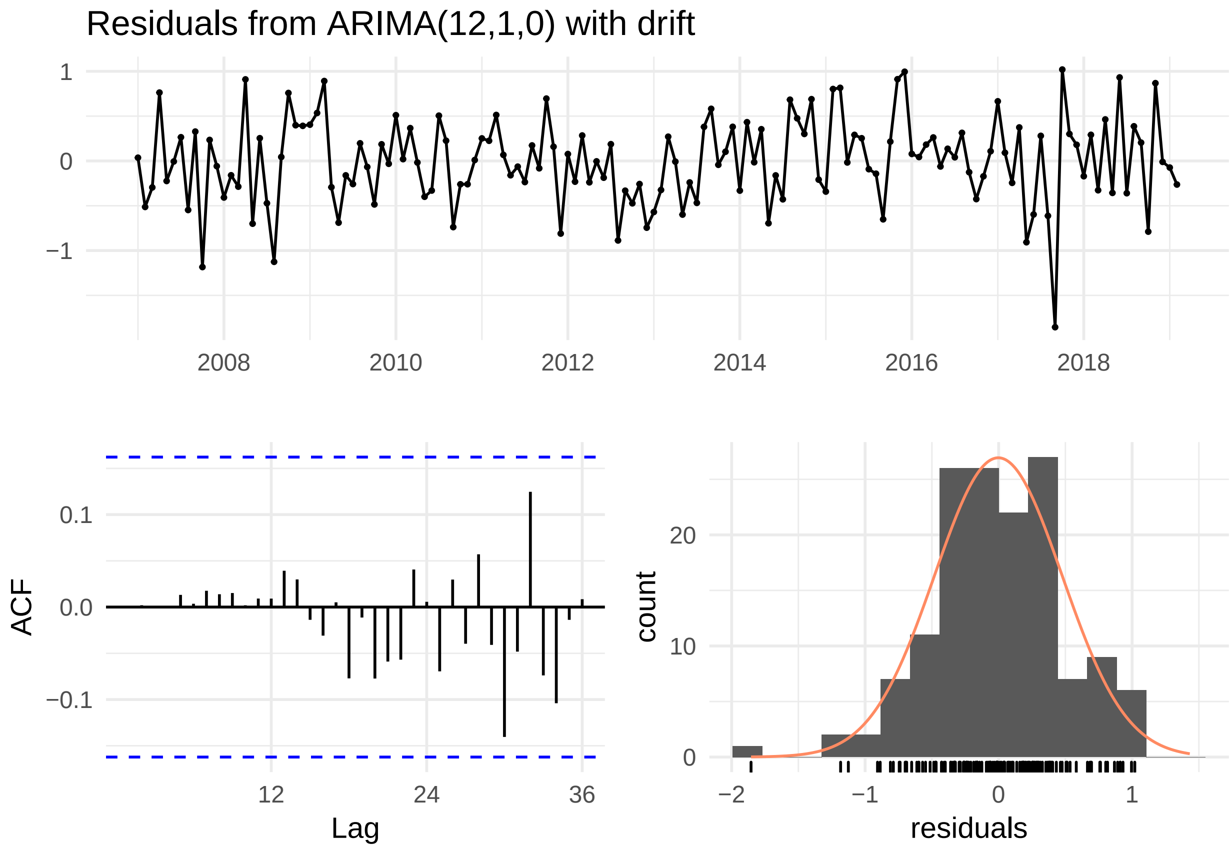
Here is a purely auto regressive model of Average weekly hours. $p = 12, \ d = 1, \ q = 0$

$p$ represents the number of lagged values to use, $d$ is the order of differencing. $q$ is the number of moving average terms.

```
## Series: TSA[, "Average_Weekly_Hours"]
## ARIMA(12,1,0) with drift
##
## Coefficients:
##           ar1      ar2      ar3      ar4      ar5      ar6      ar7
##       -0.3759  -0.3435  -0.1403  -0.0404  -0.1821  -0.2420  -0.1485
## s.e.   0.0834   0.0888   0.0933   0.0945   0.0937   0.0932   0.0945
```

```
##             ar8      ar9     ar10    ar11     ar12    drift
##         -0.1635  -0.0719  -0.1053  0.0225  -0.0212  -0.0217
## s.e.    0.0930   0.0966   0.0960  0.0912   0.0881   0.0144
##
## sigma^2 estimated as 0.2511:  log likelihood=-99.14
## AIC=226.29   AICc=229.52   BIC=267.96
##
## Training set error measures:
##                          ME      RMSE       MAE         MPE     MAPE    MASE
## Training set -0.002418786 0.4764539 0.3725031 -0.02105377 1.106984 0.52709
##                         ACF1
## Training set -0.0001733962
```





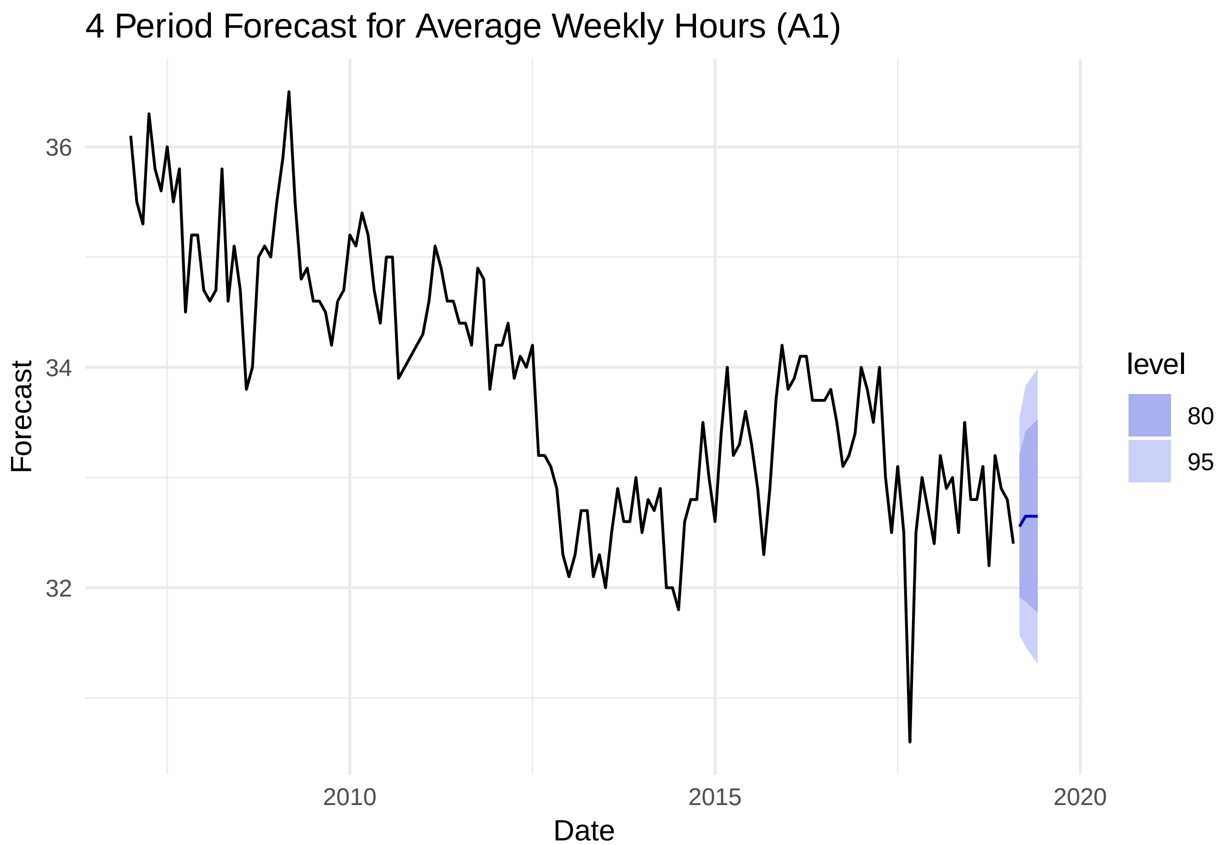

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(12,1,0) with drift
## Q* = 4.2716, df = 11, p-value = 0.9613
##
## Model df: 13.   Total lags used: 24
```

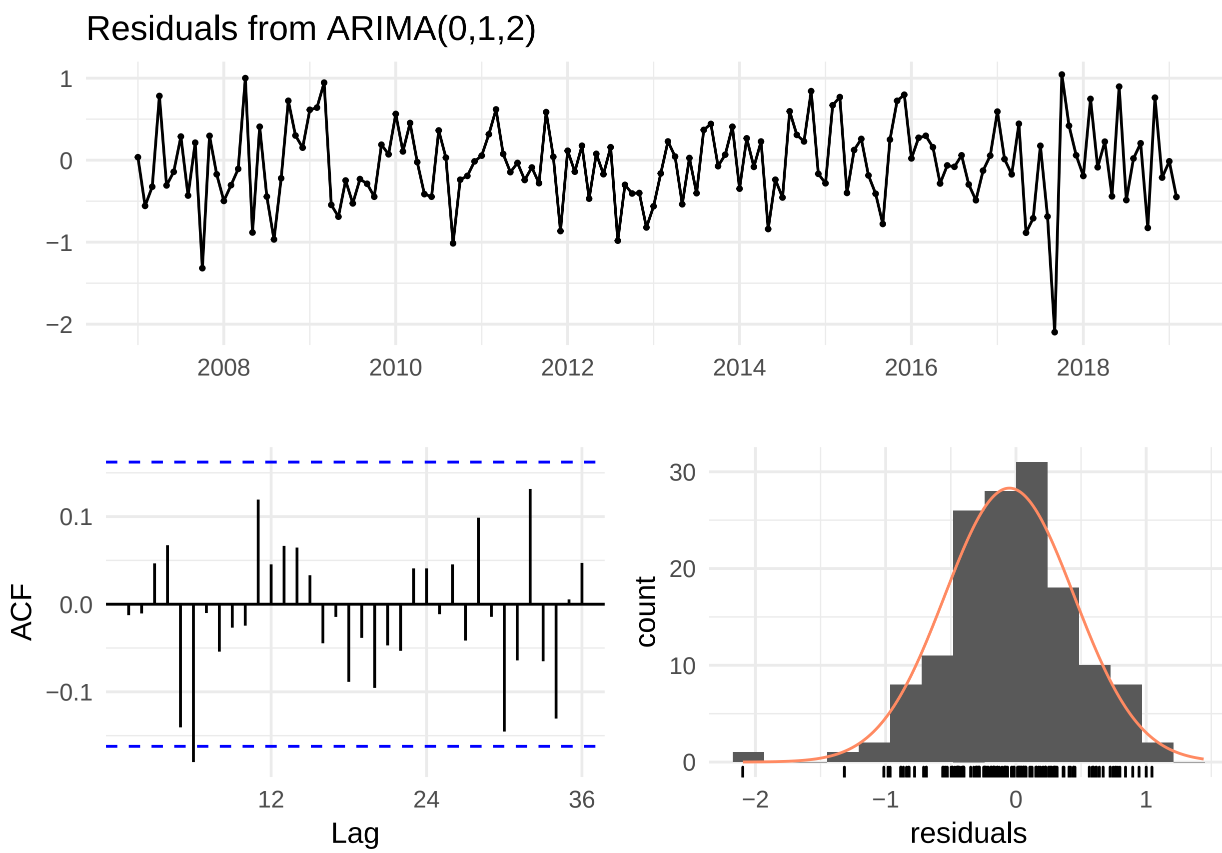
Inspecting the ACF we can see that we don't have stationarity in the data. The plot of our residuals shows some seasonal pattern so we can be sure that there is still information that is not being captured in the model. The model output shows that our out of sample root mean square error is 0.47 which is pretty good considering the simplicity of the model.

### Automatic ARIMA for Average Weekly Hours

```
## Series: TSA[, "Average_Weekly_Hours"]
## ARIMA(0,1,2)
##
## Coefficients:
##           ma1      ma2
##       -0.3399  -0.2076
## s.e.   0.0824   0.0879
##
## sigma^2 estimated as 0.2532:  log likelihood=-105.31
## AIC=216.62   AICc=216.79   BIC=225.55
```

```
##
## Training set error measures:
##                       ME      RMSE       MAE        MPE     MAPE      MASE
## Training set -0.05032925 0.4980414 0.3861159 -0.1638461 1.146615 0.5463519
##                     ACF1
## Training set -0.01245998
```






```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,2)
## Q* = 18.953, df = 22, p-value = 0.6482
##
## Model df: 2.   Total lags used: 24
```

A lack of serial correlation can be observed due the high p-value of the ljung box test. The residual graph shows an ACF where there might be dependence at some of the later lags. This model is average with a root mean squared error of 0.498

#### Variable Selection for Average Weekly Hours

```
## Forward Selection Method
## ------------------------
##
## Candidate Terms:
##
## 1 . D1_Federal_Employees
## 2 . D1_Healthcare_Employees
## 3 . D1_Local_Gov_Employees
## 4 . D1_Retail_Employees
## 5 . D1_All_Employees
```

```
## 6 . D1_Average_Weekly_Hours
## 7 . D1_Average_Hourly_Earnings
## 8 . D1_Average_Weekly_Earnings
## 9 . D1_Total_Weekly_Earnings
## 10 . L12_Total_Weekly_Earnings
## 11 . L11_Total_Weekly_Earnings
## 12 . L10_Total_Weekly_Earnings
## 13 . L9_Total_Weekly_Earnings
## 14 . L8_Total_Weekly_Earnings
## 15 . L7_Total_Weekly_Earnings
## 16 . L6_Total_Weekly_Earnings
## 17 . L5_Total_Weekly_Earnings
## 18 . L4_Total_Weekly_Earnings
## 19 . L3_Total_Weekly_Earnings
## 20 . L2_Total_Weekly_Earnings
## 21 . L1_Total_Weekly_Earnings
## 22 . L12_Average_Weekly_Earnings
## 23 . L11_Average_Weekly_Earnings
## 24 . L10_Average_Weekly_Earnings
## 25 . L9_Average_Weekly_Earnings
## 26 . L8_Average_Weekly_Earnings
## 27 . L7_Average_Weekly_Earnings
## 28 . L6_Average_Weekly_Earnings
## 29 . L5_Average_Weekly_Earnings
## 30 . L4_Average_Weekly_Earnings
## 31 . L3_Average_Weekly_Earnings
## 32 . L2_Average_Weekly_Earnings
## 33 . L1_Average_Weekly_Earnings
## 34 . L12_Average_Hourly_Earnings
## 35 . L11_Average_Hourly_Earnings
## 36 . L10_Average_Hourly_Earnings
## 37 . L9_Average_Hourly_Earnings
## 38 . L8_Average_Hourly_Earnings
## 39 . L7_Average_Hourly_Earnings
## 40 . L6_Average_Hourly_Earnings
## 41 . L5_Average_Hourly_Earnings
## 42 . L4_Average_Hourly_Earnings
## 43 . L3_Average_Hourly_Earnings
## 44 . L2_Average_Hourly_Earnings
## 45 . L1_Average_Hourly_Earnings
## 46 . L12_Average_Weekly_Hours
## 47 . L11_Average_Weekly_Hours
## 48 . L10_Average_Weekly_Hours
## 49 . L9_Average_Weekly_Hours
## 50 . L8_Average_Weekly_Hours
## 51 . L7_Average_Weekly_Hours
## 52 . L6_Average_Weekly_Hours
## 53 . L5_Average_Weekly_Hours
## 54 . L4_Average_Weekly_Hours
## 55 . L3_Average_Weekly_Hours
## 56 . L2_Average_Weekly_Hours
## 57 . L1_Average_Weekly_Hours
## 58 . L12_All_Employees
## 59 . L11_All_Employees
```

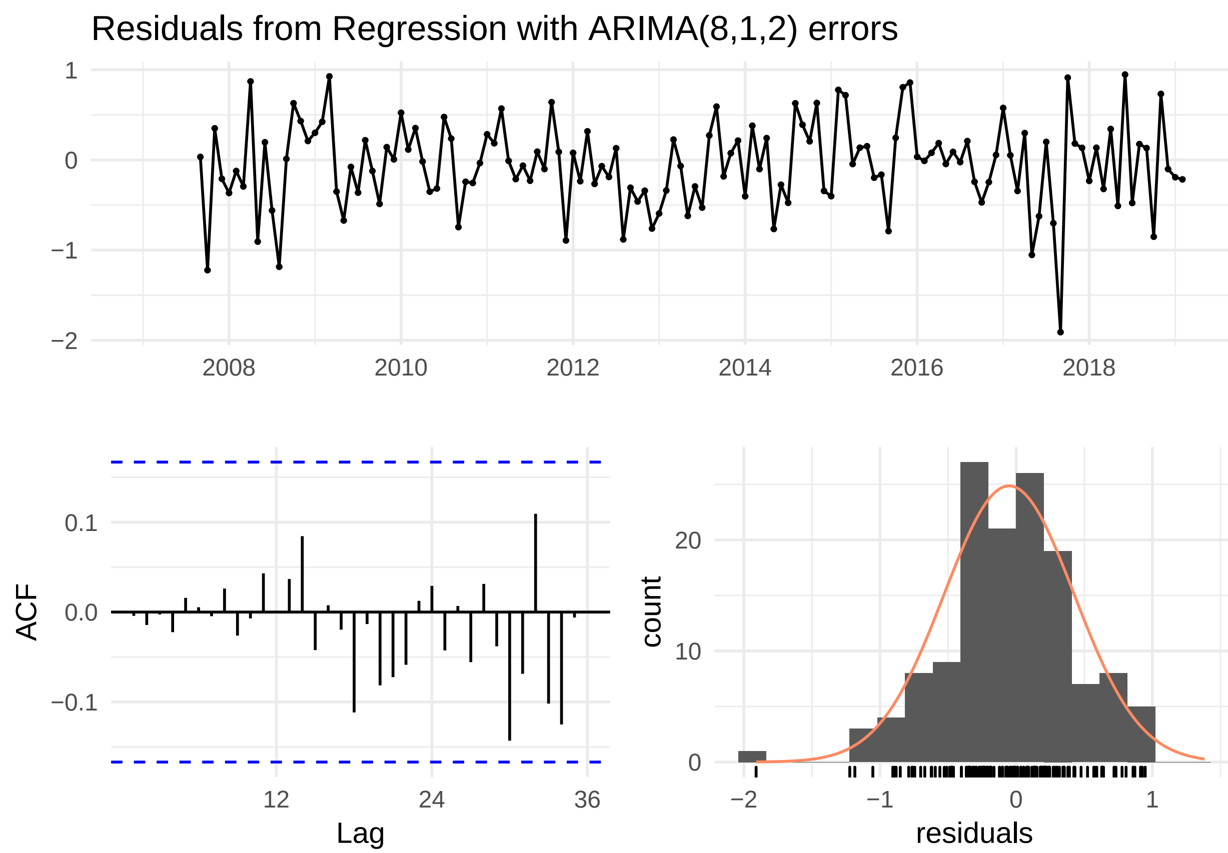
```
## 60 . L10_All_Employees
## 61 . L9_All_Employees
## 62 . L8_All_Employees
## 63 . L7_All_Employees
## 64 . L6_All_Employees
## 65 . L5_All_Employees
## 66 . L4_All_Employees
## 67 . L3_All_Employees
## 68 . L2_All_Employees
## 69 . L1_All_Employees
## 70 . L12_Retail_Employees
## 71 . L11_Retail_Employees
## 72 . L10_Retail_Employees
## 73 . L9_Retail_Employees
## 74 . L8_Retail_Employees
## 75 . L7_Retail_Employees
## 76 . L6_Retail_Employees
## 77 . L5_Retail_Employees
## 78 . L4_Retail_Employees
## 79 . L3_Retail_Employees
## 80 . L2_Retail_Employees
## 81 . L1_Retail_Employees
## 82 . L12_Local_Gov_Employees
## 83 . L11_Local_Gov_Employees
## 84 . L10_Local_Gov_Employees
## 85 . L9_Local_Gov_Employees
## 86 . L8_Local_Gov_Employees
## 87 . L7_Local_Gov_Employees
## 88 . L6_Local_Gov_Employees
## 89 . L5_Local_Gov_Employees
## 90 . L4_Local_Gov_Employees
## 91 . L3_Local_Gov_Employees
## 92 . L2_Local_Gov_Employees
## 93 . L1_Local_Gov_Employees
## 94 . L12_Healthcare_Employees
## 95 . L11_Healthcare_Employees
## 96 . L10_Healthcare_Employees
## 97 . L9_Healthcare_Employees
## 98 . L8_Healthcare_Employees
## 99 . L7_Healthcare_Employees
## 100 . L6_Healthcare_Employees
## 101 . L5_Healthcare_Employees
## 102 . L4_Healthcare_Employees
## 103 . L3_Healthcare_Employees
## 104 . L2_Healthcare_Employees
## 105 . L1_Healthcare_Employees
## 106 . L12_Federal_Employees
## 107 . L11_Federal_Employees
## 108 . L10_Federal_Employees
## 109 . L9_Federal_Employees
## 110 . L8_Federal_Employees
## 111 . L7_Federal_Employees
## 112 . L6_Federal_Employees
## 113 . L5_Federal_Employees
```

```
## 114 . L4_Federal_Employees
## 115 . L3_Federal_Employees
## 116 . L2_Federal_Employees
## 117 . L1_Federal_Employees
## 118 . Federal_Employees
## 119 . Healthcare_Employees
## 120 . Local_Gov_Employees
## 121 . Retail_Employees
## 122 . All_Employees
## 123 . Average_Hourly_Earnings
## 124 . Average_Weekly_Earnings
## 125 . Total_Weekly_Earnings
##
##
## Variables Entered:
##
## - L1_Average_Weekly_Hours
## - D1_Average_Weekly_Hours
## - L2_Local_Gov_Employees
## - D1_Total_Weekly_Earnings
## - D1_Average_Weekly_Earnings
##
## No more variables to be added.
```

```
##
##                                  Selection Summary
## ------------------------------------------------------------------------------------
## Variable                          AIC       Sum Sq      RSS       R-Sq      Adj. R-Sq
## ------------------------------------------------------------------------------------
## L1_Average_Weekly_Hours         227.678    141.021    39.161    0.78266      0.78114
## D1_Average_Weekly_Hours       -9101.524    180.182     0.000    1.00000      1.00000
## L2_Local_Gov_Employees        -9379.775    177.297     0.000    1.00000      1.00000
## D1_Total_Weekly_Earnings      -9384.215    177.297     0.000    1.00000      1.00000
## D1_Average_Weekly_Earnings    -9384.887    177.297     0.000    1.00000      1.00000
## ------------------------------------------------------------------------------------
```

The step-wise regression selected a large number of $x$ variables. Based on intuition this is likely to overfit so I'll be using the most sensiscal variables it selected for the final model. This will also be based on the AIC of the variables.

- Lags 1,3,8 of Average Hourly Earnings, Average Weekly Earnings, Local Government Employment and Federal Government Employment
- Lags 4,8 Healthcare employment and Retail employment
- Difference Average Weekly Hours once
- Use 8 lags of the dependent variable
- Use 2 moving average terms (based on automatic arima evaluation for purely auto regressive model above)




```
##
##  Ljung-Box test
##
## data:  Residuals from Regression with ARIMA(8,1,2) errors
## Q* = 7.0444, df = 13, p-value = 0.8999
##
## Model df: 11.   Total lags used: 24
```

```
## Series: TSA[, "Average_Weekly_Hours"]
## Regression with ARIMA(8,1,2) errors
##
## Coefficients:
##           ar1     ar2      ar3     ar4      ar5      ar6      ar7      ar8
##       -0.6130  0.1044  -0.0352  0.1249  -0.1398  -0.2771  -0.0937  -0.0918
## s.e.   0.2661  0.2927   0.1632  0.1255   0.1013   0.1090   0.1391   0.1110
##          ma1     ma2   xreg
##       0.2411  -0.534  9e-04
## s.e.  0.2647   0.227  1e-03
##
## sigma^2 estimated as 0.2517:  log likelihood=-94.59
## AIC=213.18   AICc=215.7   BIC=248.22
##
## Training set error measures:
##                       ME      RMSE       MAE        MPE     MAPE     MASE
## Training set -0.05222861 0.4793686 0.3692353 -0.1711456 1.102112 0.522466
```

```
##                      ACF1
## Training set -0.00418556
```

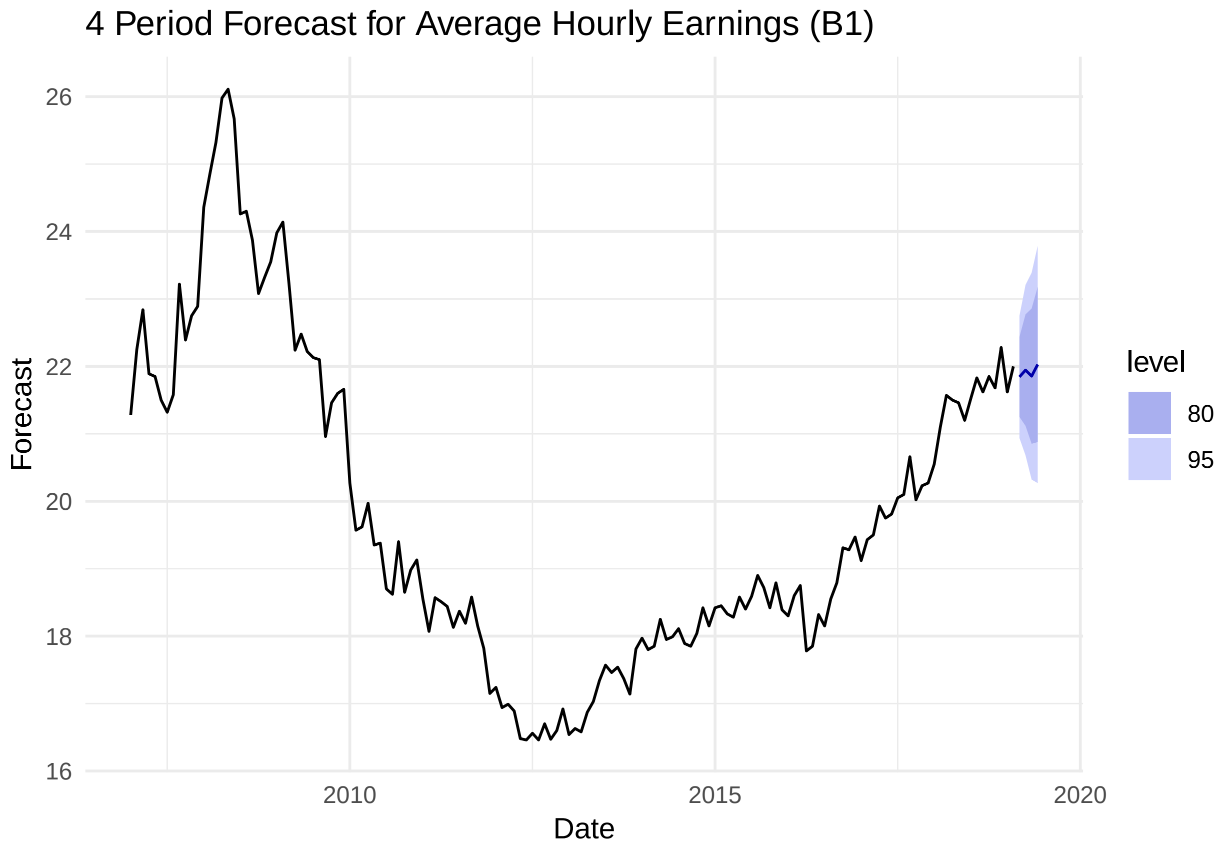
```
## # A tibble: 3 x 6
##   `Point Forecast` `Lo 80` `Hi 80` `Lo 95` `Hi 95` Date
##              <dbl>   <dbl>   <dbl>   <dbl>   <dbl> <chr>
## 1             32.9    31.9    33.9    31.4    34.5 March
## 2             33.1    32.0    34.1    31.5    34.6 April
## 3             33.0    31.9    34.0    31.4    34.5 May
```

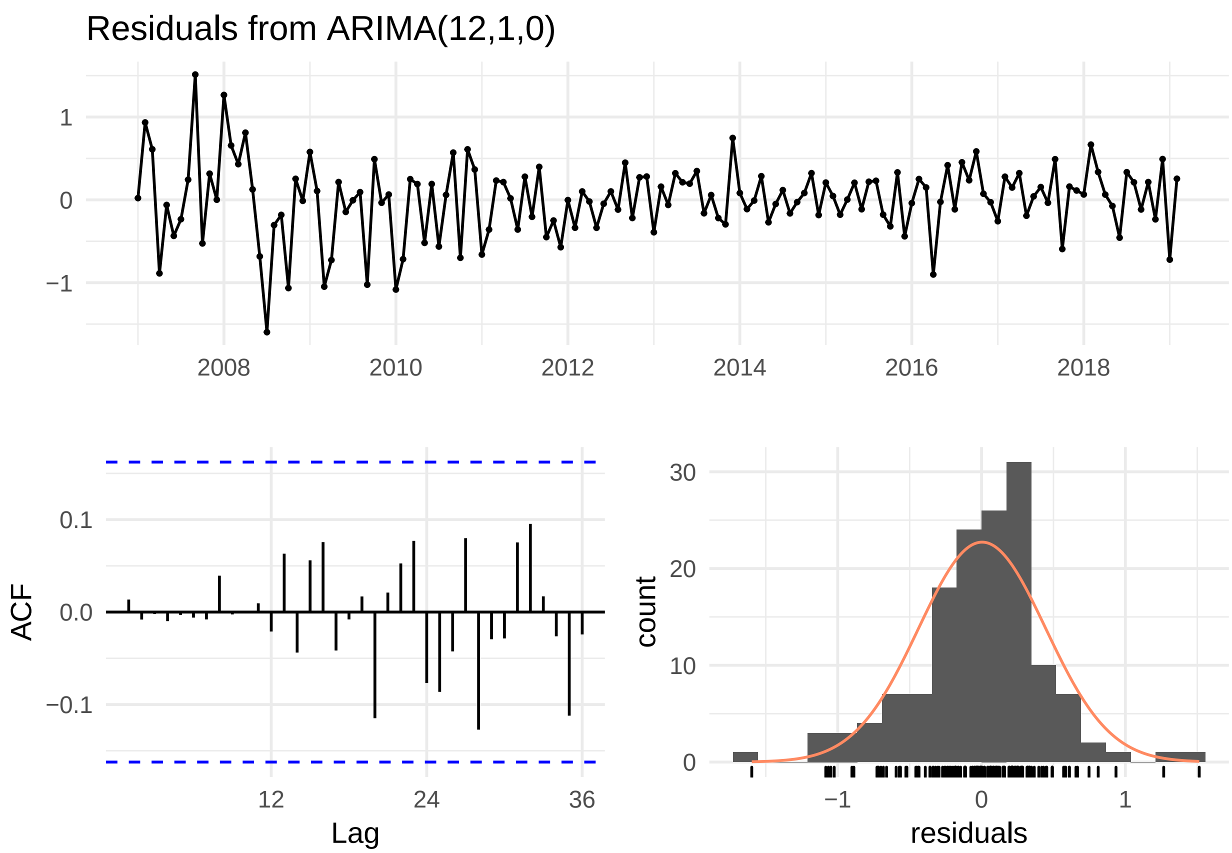
The test for serial correlation was near a p-value of 0.9 which suggest a large lack of evidence for serial correlation. The model itself has a low root mean squared error. The only concern is the possibility of a seasonal trend within the plot of the residuals.

## Average Hourly Earnings

### ARIMA of Average Hourly Earnings (12,1,0)

```
## Series: TSA[, "Average_Hourly_Earnings"]
## ARIMA(12,1,0)
##
## Coefficients:
##           ar1      ar2      ar3     ar4      ar5     ar6      ar7      ar8
##       -0.0272  -0.0158  -0.0066  0.1041  -0.0488  0.1178  -0.0081  -0.0851
## s.e.   0.0835   0.0836   0.0854  0.0847   0.0847  0.0849   0.0848   0.0884
##          ar9    ar10    ar11     ar12
##       0.1160  0.0397  0.1384  -0.0585
## s.e.  0.0893  0.0896  0.0892   0.0937
##
## sigma^2 estimated as 0.2139:  log likelihood=-87.95
## AIC=201.9   AICc=204.68   BIC=240.6
##
## Training set error measures:
##                       ME      RMSE       MAE         MPE     MAPE
## Training set 0.003927872 0.4413727 0.3251383 0.004430914 1.607209
##                   MASE       ACF1
## Training set 0.2297797 0.01353898
```


4 Period Forecast for Average Hourly Earnings (B1)
Forecast
26
24
22
20
18
16
2010
2015
2020
Date
level
80
95



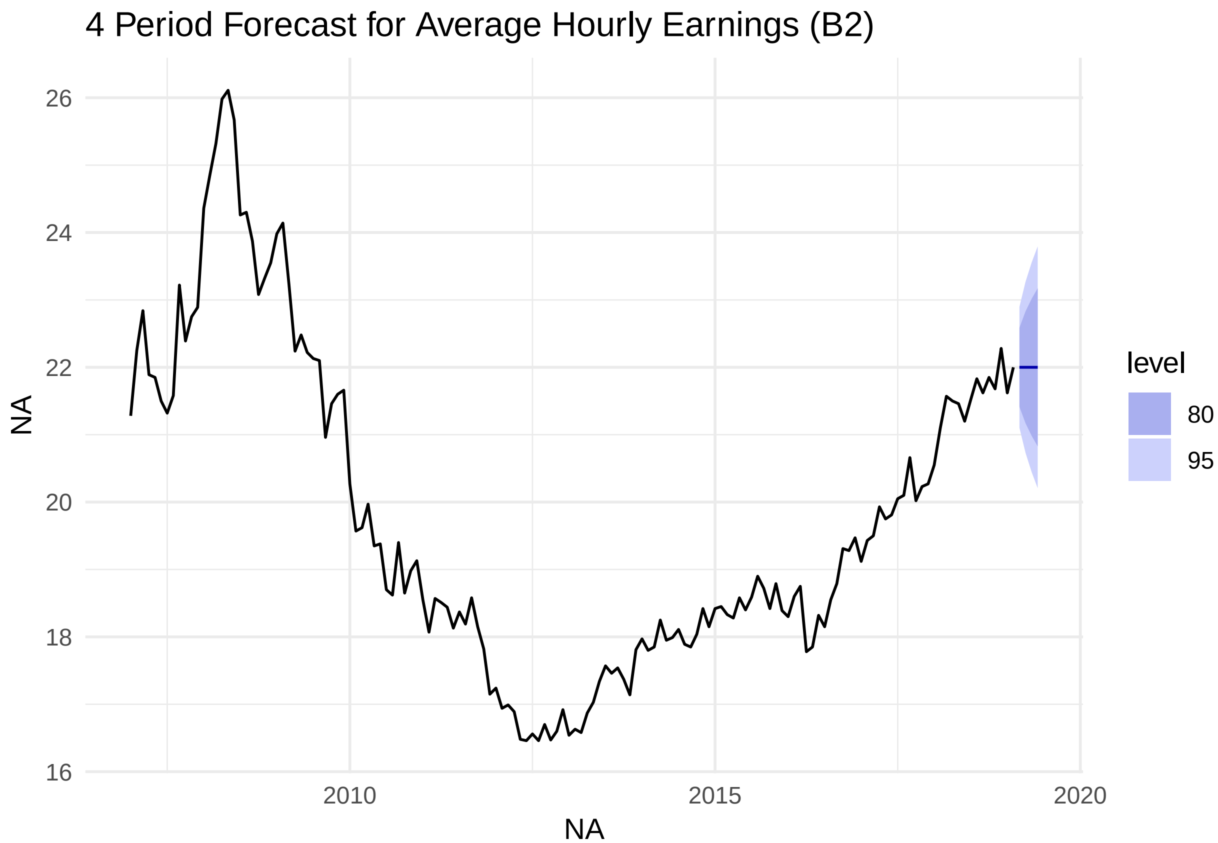

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(12,1,0)
## Q* = 8.0818, df = 12, p-value = 0.7787
##
## Model df: 12.   Total lags used: 24
```

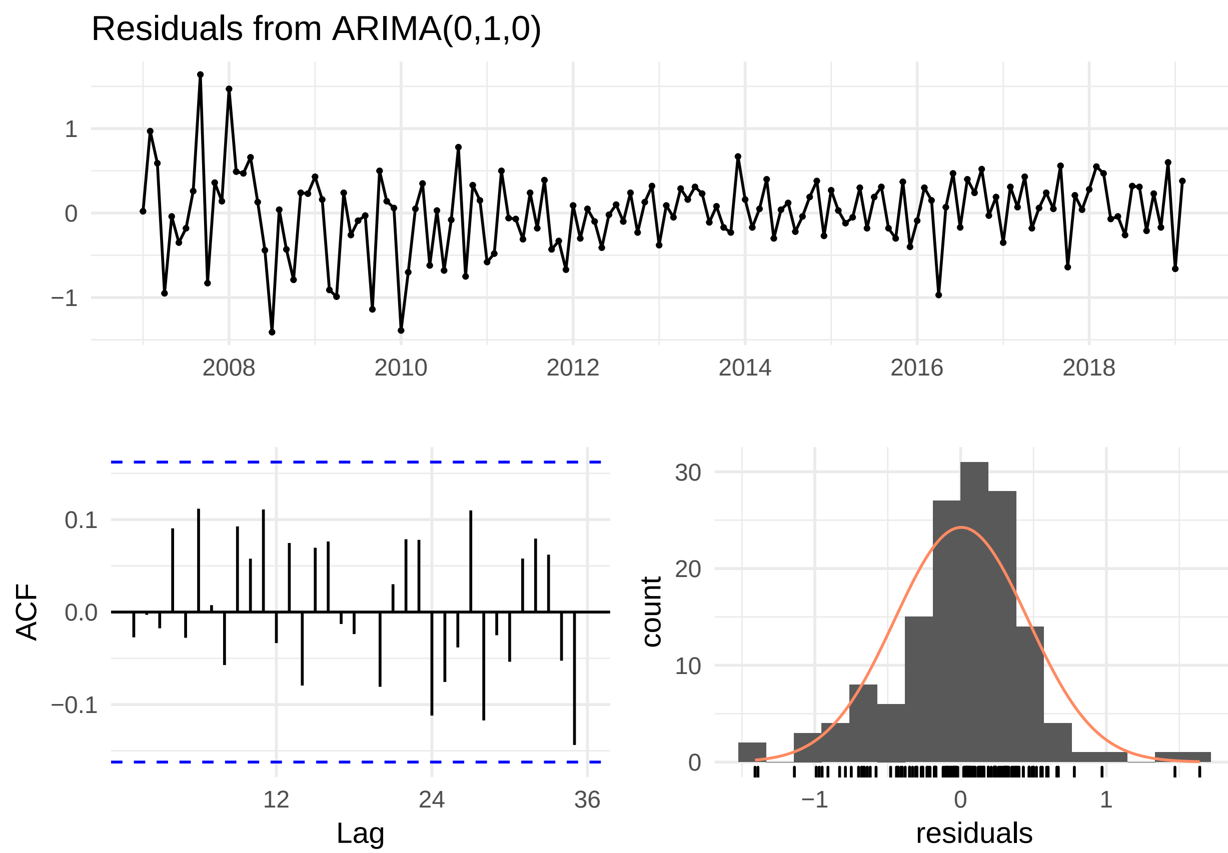
This model had a surprisingly good outcome considering it was being used as a baseline for this project. The Ljung Box tests suggests no serial correlation, or rather the lack there of. The root mean squared error is fairly low and a plot of the residuals is untelling of missed opportunities to extract information from our residuals.

#### Automatic ARIMA for Average Hourly Earnings

```
## Series: TSA[, "Average_Hourly_Earnings"]
## ARIMA(0,1,0)
##
## sigma^2 estimated as 0.2096:  log likelihood=-92.46
## AIC=186.93   AICc=186.96   BIC=189.91
##
## Training set error measures:
##                      ME      RMSE       MAE           MPE     MAPE
## Training set 0.00507726 0.4562658 0.3385019 -0.0003650678 1.675021
##                   MASE        ACF1
```

```
## Training set 0.239224 -0.02729854
```






```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,0)
## Q* = 17.495, df = 24, p-value = 0.8268
##
## Model df: 0.   Total lags used: 24
```

Our Ljung box test returned a P-value of 0.8268 which shows an overwhelming lack of evidence for independence in the model. We can be cautiously confident that we don't have serial correlation in this model and it is acceptable. The root mean squared error is relatively low at 0.4562. The plot of the residuals still shows a cyclical pattern which can either be interpreted as noise or there might still be information in our residuals that we can utilize to improve this model.

### Variable Selection for Average Hourly Earnings

```
## Forward Selection Method
## ------------------------
##
## Candidate Terms:
##
## 1 . D1_Federal_Employees
## 2 . D1_Healthcare_Employees
## 3 . D1_Local_Gov_Employees
```

```
## 4 . D1_Retail_Employees
## 5 . D1_All_Employees
## 6 . D1_Average_Weekly_Hours
## 7 . D1_Average_Hourly_Earnings
## 8 . D1_Average_Weekly_Earnings
## 9 . D1_Total_Weekly_Earnings
## 10 . L12_Total_Weekly_Earnings
## 11 . L11_Total_Weekly_Earnings
## 12 . L10_Total_Weekly_Earnings
## 13 . L9_Total_Weekly_Earnings
## 14 . L8_Total_Weekly_Earnings
## 15 . L7_Total_Weekly_Earnings
## 16 . L6_Total_Weekly_Earnings
## 17 . L5_Total_Weekly_Earnings
## 18 . L4_Total_Weekly_Earnings
## 19 . L3_Total_Weekly_Earnings
## 20 . L2_Total_Weekly_Earnings
## 21 . L1_Total_Weekly_Earnings
## 22 . L12_Average_Weekly_Earnings
## 23 . L11_Average_Weekly_Earnings
## 24 . L10_Average_Weekly_Earnings
## 25 . L9_Average_Weekly_Earnings
## 26 . L8_Average_Weekly_Earnings
## 27 . L7_Average_Weekly_Earnings
## 28 . L6_Average_Weekly_Earnings
## 29 . L5_Average_Weekly_Earnings
## 30 . L4_Average_Weekly_Earnings
## 31 . L3_Average_Weekly_Earnings
## 32 . L2_Average_Weekly_Earnings
## 33 . L1_Average_Weekly_Earnings
## 34 . L12_Average_Hourly_Earnings
## 35 . L11_Average_Hourly_Earnings
## 36 . L10_Average_Hourly_Earnings
## 37 . L9_Average_Hourly_Earnings
## 38 . L8_Average_Hourly_Earnings
## 39 . L7_Average_Hourly_Earnings
## 40 . L6_Average_Hourly_Earnings
## 41 . L5_Average_Hourly_Earnings
## 42 . L4_Average_Hourly_Earnings
## 43 . L3_Average_Hourly_Earnings
## 44 . L2_Average_Hourly_Earnings
## 45 . L1_Average_Hourly_Earnings
## 46 . L12_Average_Weekly_Hours
## 47 . L11_Average_Weekly_Hours
## 48 . L10_Average_Weekly_Hours
## 49 . L9_Average_Weekly_Hours
## 50 . L8_Average_Weekly_Hours
## 51 . L7_Average_Weekly_Hours
## 52 . L6_Average_Weekly_Hours
## 53 . L5_Average_Weekly_Hours
## 54 . L4_Average_Weekly_Hours
## 55 . L3_Average_Weekly_Hours
## 56 . L2_Average_Weekly_Hours
## 57 . L1_Average_Weekly_Hours
```

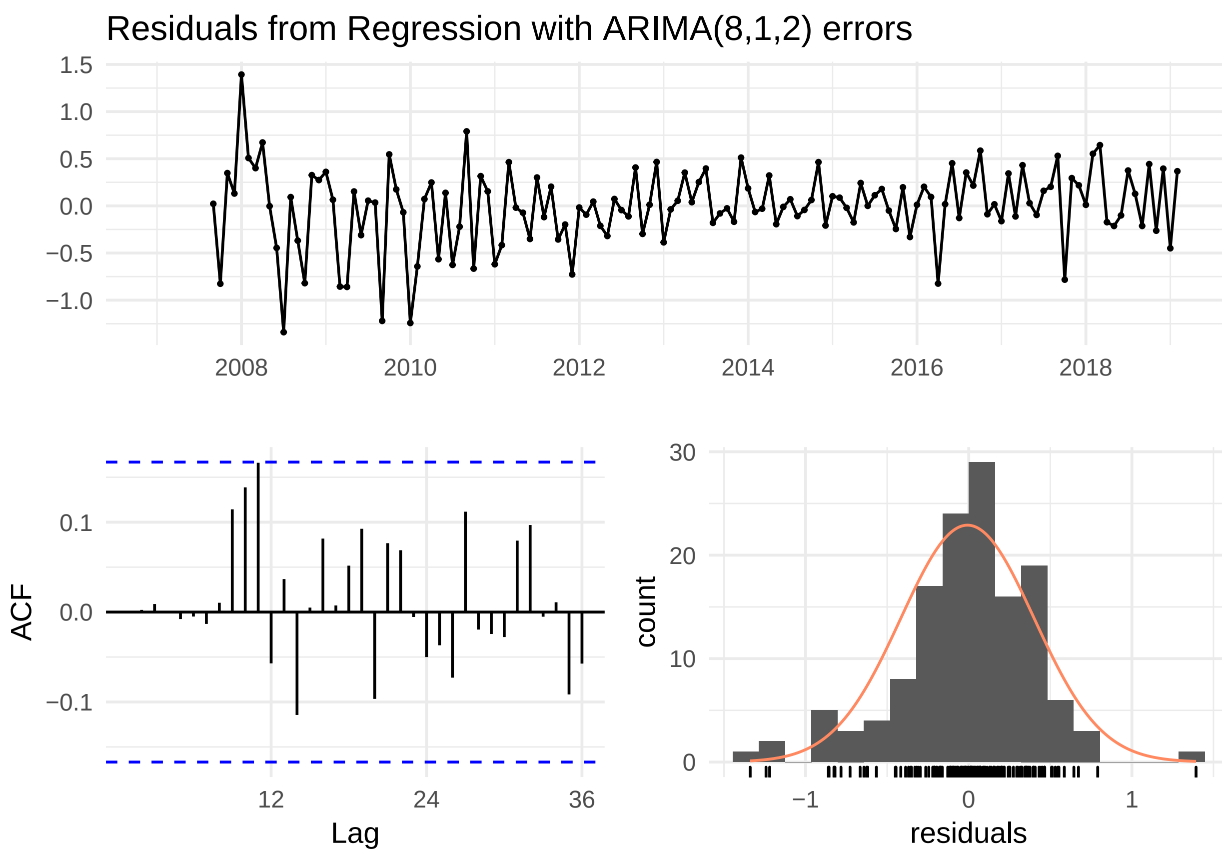
```
## 58 . L12_All_Employees
## 59 . L11_All_Employees
## 60 . L10_All_Employees
## 61 . L9_All_Employees
## 62 . L8_All_Employees
## 63 . L7_All_Employees
## 64 . L6_All_Employees
## 65 . L5_All_Employees
## 66 . L4_All_Employees
## 67 . L3_All_Employees
## 68 . L2_All_Employees
## 69 . L1_All_Employees
## 70 . L12_Retail_Employees
## 71 . L11_Retail_Employees
## 72 . L10_Retail_Employees
## 73 . L9_Retail_Employees
## 74 . L8_Retail_Employees
## 75 . L7_Retail_Employees
## 76 . L6_Retail_Employees
## 77 . L5_Retail_Employees
## 78 . L4_Retail_Employees
## 79 . L3_Retail_Employees
## 80 . L2_Retail_Employees
## 81 . L1_Retail_Employees
## 82 . L12_Local_Gov_Employees
## 83 . L11_Local_Gov_Employees
## 84 . L10_Local_Gov_Employees
## 85 . L9_Local_Gov_Employees
## 86 . L8_Local_Gov_Employees
## 87 . L7_Local_Gov_Employees
## 88 . L6_Local_Gov_Employees
## 89 . L5_Local_Gov_Employees
## 90 . L4_Local_Gov_Employees
## 91 . L3_Local_Gov_Employees
## 92 . L2_Local_Gov_Employees
## 93 . L1_Local_Gov_Employees
## 94 . L12_Healthcare_Employees
## 95 . L11_Healthcare_Employees
## 96 . L10_Healthcare_Employees
## 97 . L9_Healthcare_Employees
## 98 . L8_Healthcare_Employees
## 99 . L7_Healthcare_Employees
## 100 . L6_Healthcare_Employees
## 101 . L5_Healthcare_Employees
## 102 . L4_Healthcare_Employees
## 103 . L3_Healthcare_Employees
## 104 . L2_Healthcare_Employees
## 105 . L1_Healthcare_Employees
## 106 . L12_Federal_Employees
## 107 . L11_Federal_Employees
## 108 . L10_Federal_Employees
## 109 . L9_Federal_Employees
## 110 . L8_Federal_Employees
## 111 . L7_Federal_Employees
```

```
## 112 . L6_Federal_Employees
## 113 . L5_Federal_Employees
## 114 . L4_Federal_Employees
## 115 . L3_Federal_Employees
## 116 . L2_Federal_Employees
## 117 . L1_Federal_Employees
## 118 . Federal_Employees
## 119 . Healthcare_Employees
## 120 . Local_Gov_Employees
## 121 . Retail_Employees
## 122 . All_Employees
## 123 . Average_Hourly_Earnings
## 124 . Average_Weekly_Earnings
## 125 . Total_Weekly_Earnings
##
##
## Variables Entered:
##
## - L1_Average_Weekly_Hours
## - D1_Average_Weekly_Hours
## - L2_Local_Gov_Employees
## - D1_Total_Weekly_Earnings
## - D1_Average_Weekly_Earnings
##
## No more variables to be added.
```

```
##
##                                   Selection Summary
## ---------------------------------------------------------------------------------------
## Variable                        AIC        Sum Sq      RSS       R-Sq      Adj. R-Sq
## ---------------------------------------------------------------------------------------
## L1_Average_Weekly_Hours         227.678    141.021    39.161    0.78266      0.78114
## D1_Average_Weekly_Hours       -9101.524    180.182     0.000    1.00000      1.00000
## L2_Local_Gov_Employees        -9379.775    177.297     0.000    1.00000      1.00000
## D1_Total_Weekly_Earnings      -9384.215    177.297     0.000    1.00000      1.00000
## D1_Average_Weekly_Earnings    -9384.887    177.297     0.000    1.00000      1.00000
## ---------------------------------------------------------------------------------------
```

Based on the guidance of the stepwise regression. Using intuition, PACs and gut feel I'll be building an model with the following variables.

- Lag 6 of Local and Federal employment
- 8 Lags of Average Hourly Earnings
- 2 Moving average terms
- 1 order differencing
- Lags 1,3,8 of Average Weekly Earnings




```
##
##  Ljung-Box test
##
## data:  Residuals from Regression with ARIMA(8,1,2) errors
## Q* = 18.484, df = 13, p-value = 0.14
##
## Model df: 11.   Total lags used: 24

## Series: TSA[, "Average_Hourly_Earnings"]
## Regression with ARIMA(8,1,2) errors
##
## Coefficients:
##           ar1      ar2      ar3     ar4      ar5     ar6      ar7      ar8
##       -0.5396  -0.8983  -0.0325  0.0407  -0.0179  0.0567  -0.0304  -0.0595
## s.e.   0.1454   0.1071   0.2255  0.1492   0.1535  0.1441   0.1068   0.1035
##          ma1     ma2    xreg
##       0.5506  1.0000  0.0004
## s.e.  0.0250  0.0404  0.0024
##
## sigma^2 estimated as 0.1818:  log likelihood=-74.07
## AIC=172.15   AICc=174.67   BIC=207.19
##
## Training set error measures:
##                        ME      RMSE       MAE         MPE     MAPE
## Training set -0.007677987 0.4074091 0.2969296 -0.05258859 1.479158
```

```
##                   MASE         ACF1
## Training set 0.2098442 -0.001491287
```

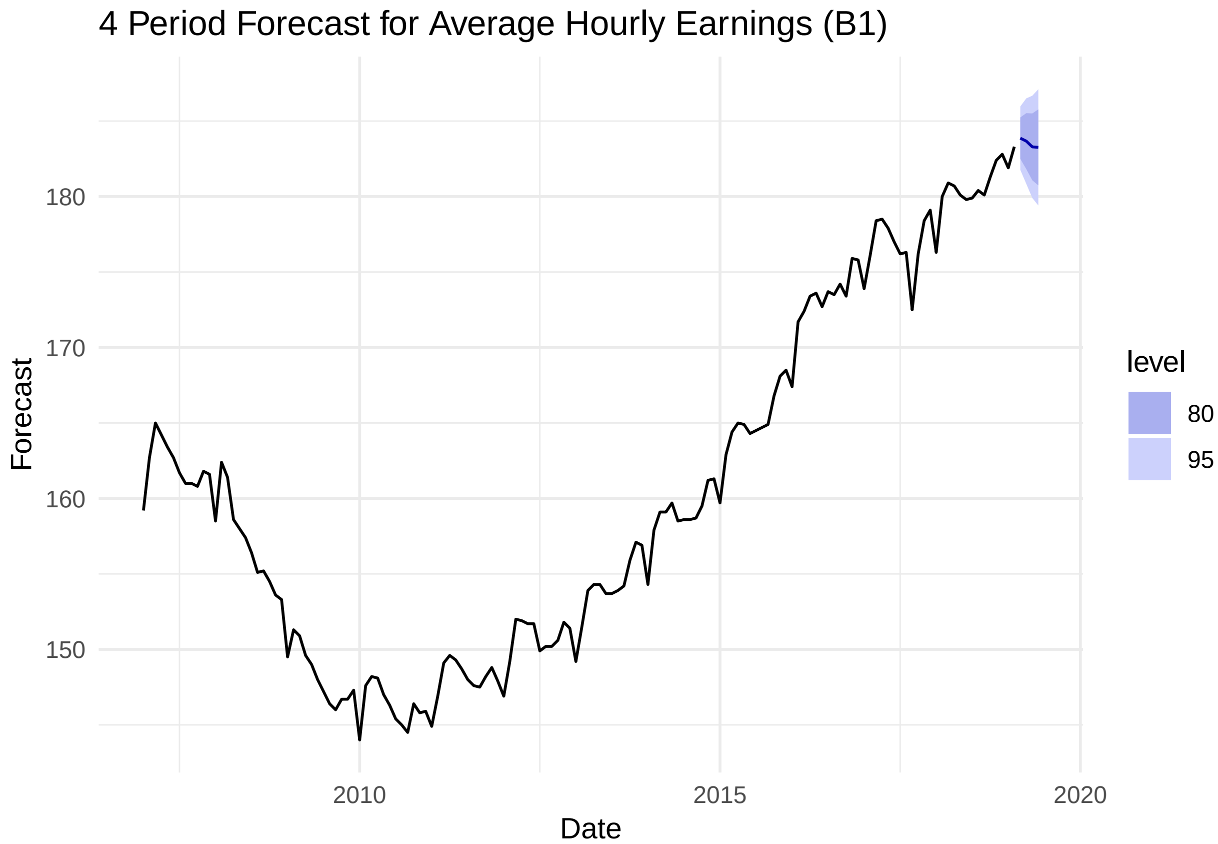
```
## # A tibble: 3 x 6
##   `Point Forecast` `Lo 80` `Hi 80` `Lo 95` `Hi 95` Date
##              <dbl>   <dbl>   <dbl>   <dbl>   <dbl> <chr>
## 1             22.1    20.4    23.8    19.5    24.7 March
## 2             22.1    20.3    23.9    19.4    24.9 April
## 3             22.2    20.3    24.1    19.3    25.1 May
```

The Ljung box test shows us that there is a lack of evidence for dependence in our variables, we have defeated serial correlation to our knowledge. The plot of the residuals seems to be somewhat cyclical but it can be hard to tell. This is an acceptable model. Our out of sample root mean squared error is 0.407 which I would consider quite good considering the scale of our predictions.

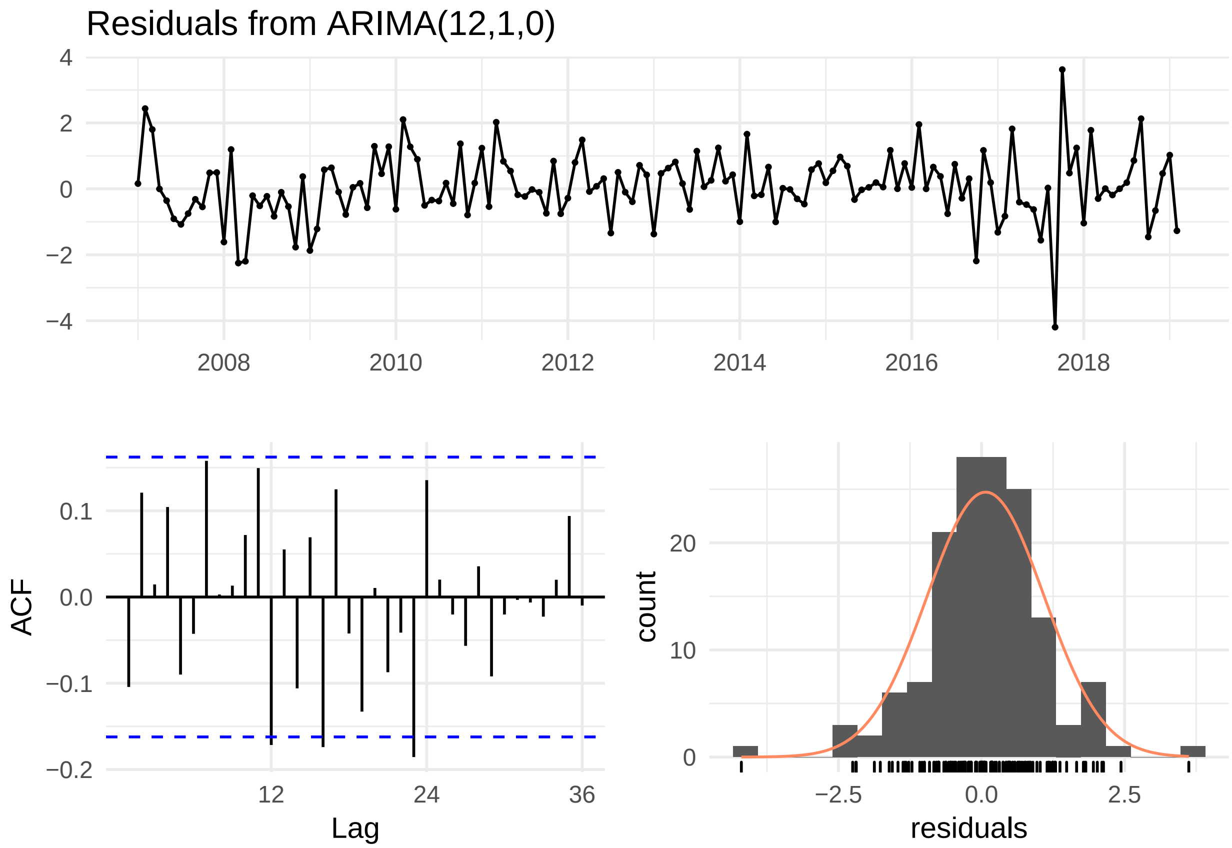
## Total Employment

### ARIMA of Total Employment (12,1,0)

```
## Series: TSA[, "All_Employees"]
## ARIMA(12,1,0)
##
## Coefficients:
##           ar1      ar2      ar3     ar4      ar5     ar6     ar7     ar8
##       -0.0941  -0.0306  -0.0156  0.0359  -0.0164  0.0288  0.0589  0.0316
## s.e.   0.0595   0.0602   0.0599  0.0599   0.0601  0.0567  0.0595  0.0600
##          ar9     ar10    ar11    ar12
##       0.0356  -0.0701  0.0524  0.6807
## s.e.  0.0596   0.0596  0.0595  0.0574
##
## sigma^2 estimated as 1.147:  log likelihood=-213.45
## AIC=452.9   AICc=455.68   BIC=491.6
##
## Training set error measures:
##                      ME     RMSE       MAE        MPE      MAPE      MASE
## Training set 0.07145814 1.022179 0.7520645 0.04453877 0.4672124 0.1638378
##                    ACF1
## Training set -0.1042469
```



4 Period Forecast for Average Hourly Earnings (B1)
Forecast
180
170
160
150
2010
2015
2020
Date
level
80
95



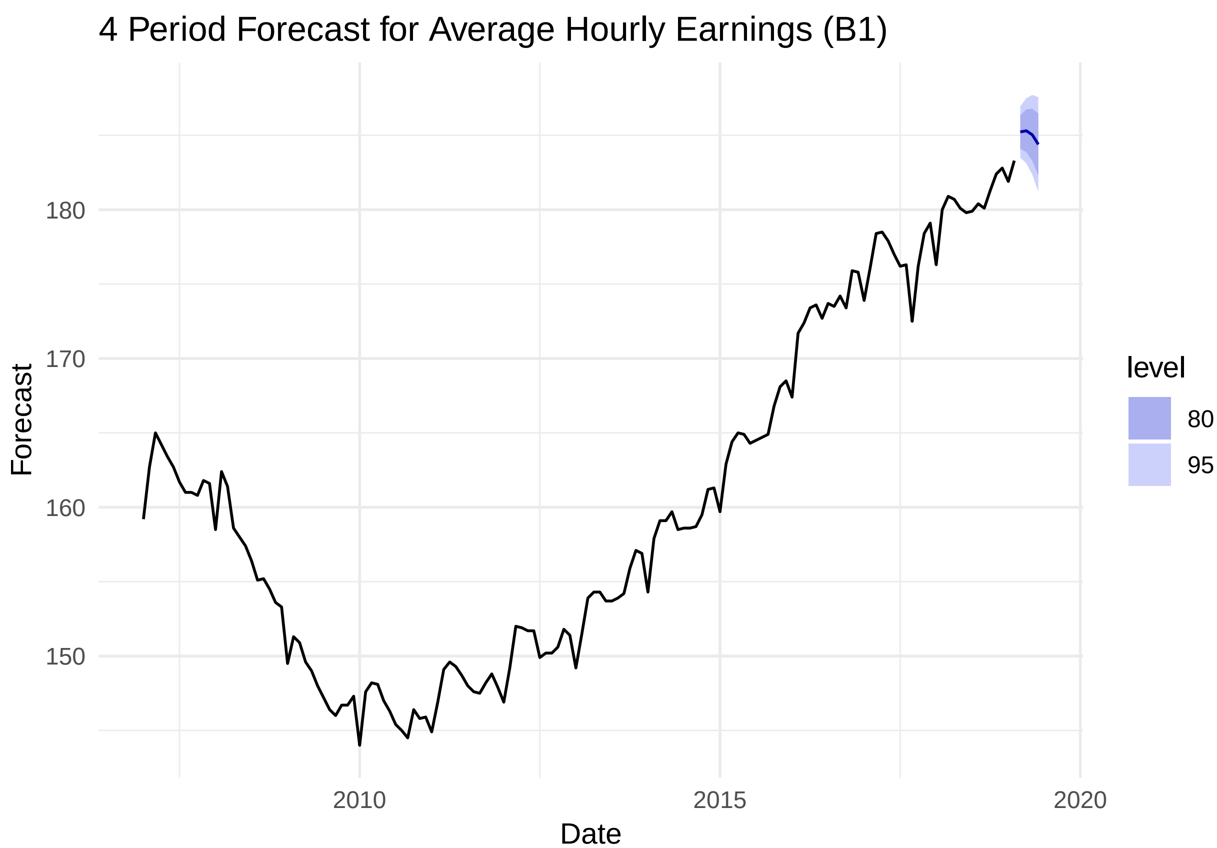

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(12,1,0)
## Q* = 45.07, df = 12, p-value = 1.003e-05
##
## Model df: 12.   Total lags used: 24
```

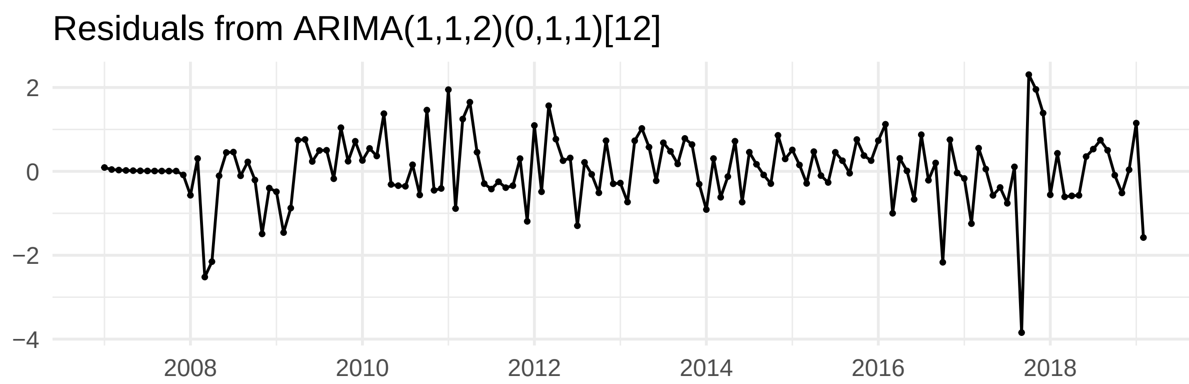
Our Ljung Box test assures us a high probability of variable dependence. This model is not dynamically complete and therefore useless.

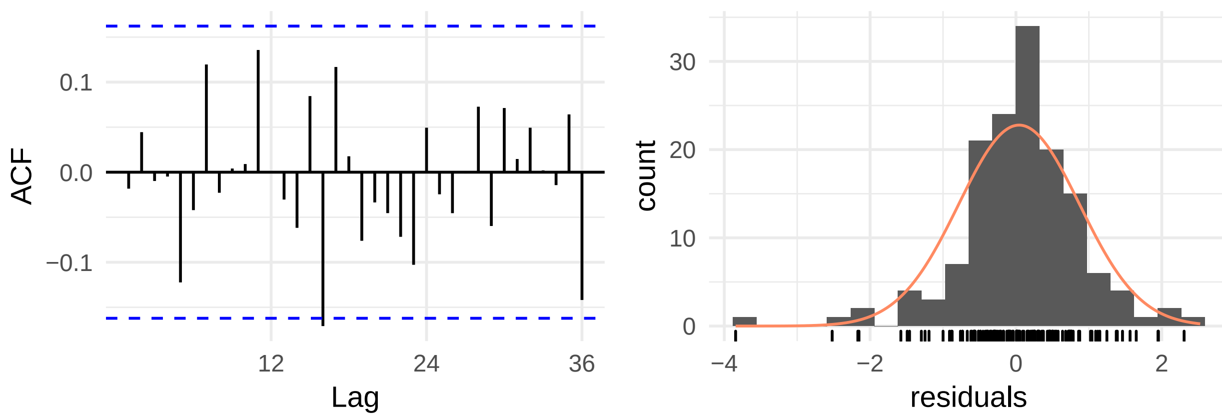
#### Automatic ARIMA for Total Employment

```
## Series: TSA[, "All_Employees"]
## ARIMA(1,1,2)(0,1,1)[12]
##
## Coefficients:
##          ar1      ma1     ma2     sma1
##       0.9553  -1.1704  0.3133  -0.7903
## s.e.  0.0396   0.0904  0.0805   0.0931
##
## sigma^2 estimated as 0.7726:  log likelihood=-174.91
## AIC=359.81   AICc=360.28   BIC=374.26
##
## Training set error measures:
```

```
##                         ME      RMSE       MAE        MPE      MAPE      MASE
## Training set 0.04349728 0.8262183 0.5924564 0.02962684 0.3702249 0.1290671
##                      ACF1
## Training set -0.0182926
```






```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,2)(0,1,1)[12]
## Q* = 22.068, df = 20, p-value = 0.3368
##
## Model df: 4.   Total lags used: 24
```

The Ljung Box test resulted in a P-Value of 0.3368 which means that there is no evidence against stationarity for this model. Examining the residual graph we can also confirm that the there is a good change we have eliminated serial correlation. Due to the complex nature of this model we can expect a dynamic model to be tricky in construction.

#### Variable Selection for Total Employment

```
## Forward Selection Method
## ------------------------
##
## Candidate Terms:
##
## 1 . D1_Federal_Employees
## 2 . D1_Healthcare_Employees
## 3 . D1_Local_Gov_Employees
## 4 . D1_Retail_Employees
```

```
## 5 . D1_All_Employees
## 6 . D1_Average_Weekly_Hours
## 7 . D1_Average_Hourly_Earnings
## 8 . D1_Average_Weekly_Earnings
## 9 . D1_Total_Weekly_Earnings
## 10 . L12_Total_Weekly_Earnings
## 11 . L11_Total_Weekly_Earnings
## 12 . L10_Total_Weekly_Earnings
## 13 . L9_Total_Weekly_Earnings
## 14 . L8_Total_Weekly_Earnings
## 15 . L7_Total_Weekly_Earnings
## 16 . L6_Total_Weekly_Earnings
## 17 . L5_Total_Weekly_Earnings
## 18 . L4_Total_Weekly_Earnings
## 19 . L3_Total_Weekly_Earnings
## 20 . L2_Total_Weekly_Earnings
## 21 . L1_Total_Weekly_Earnings
## 22 . L12_Average_Weekly_Earnings
## 23 . L11_Average_Weekly_Earnings
## 24 . L10_Average_Weekly_Earnings
## 25 . L9_Average_Weekly_Earnings
## 26 . L8_Average_Weekly_Earnings
## 27 . L7_Average_Weekly_Earnings
## 28 . L6_Average_Weekly_Earnings
## 29 . L5_Average_Weekly_Earnings
## 30 . L4_Average_Weekly_Earnings
## 31 . L3_Average_Weekly_Earnings
## 32 . L2_Average_Weekly_Earnings
## 33 . L1_Average_Weekly_Earnings
## 34 . L12_Average_Hourly_Earnings
## 35 . L11_Average_Hourly_Earnings
## 36 . L10_Average_Hourly_Earnings
## 37 . L9_Average_Hourly_Earnings
## 38 . L8_Average_Hourly_Earnings
## 39 . L7_Average_Hourly_Earnings
## 40 . L6_Average_Hourly_Earnings
## 41 . L5_Average_Hourly_Earnings
## 42 . L4_Average_Hourly_Earnings
## 43 . L3_Average_Hourly_Earnings
## 44 . L2_Average_Hourly_Earnings
## 45 . L1_Average_Hourly_Earnings
## 46 . L12_Average_Weekly_Hours
## 47 . L11_Average_Weekly_Hours
## 48 . L10_Average_Weekly_Hours
## 49 . L9_Average_Weekly_Hours
## 50 . L8_Average_Weekly_Hours
## 51 . L7_Average_Weekly_Hours
## 52 . L6_Average_Weekly_Hours
## 53 . L5_Average_Weekly_Hours
## 54 . L4_Average_Weekly_Hours
## 55 . L3_Average_Weekly_Hours
## 56 . L2_Average_Weekly_Hours
## 57 . L1_Average_Weekly_Hours
## 58 . L12_All_Employees
```

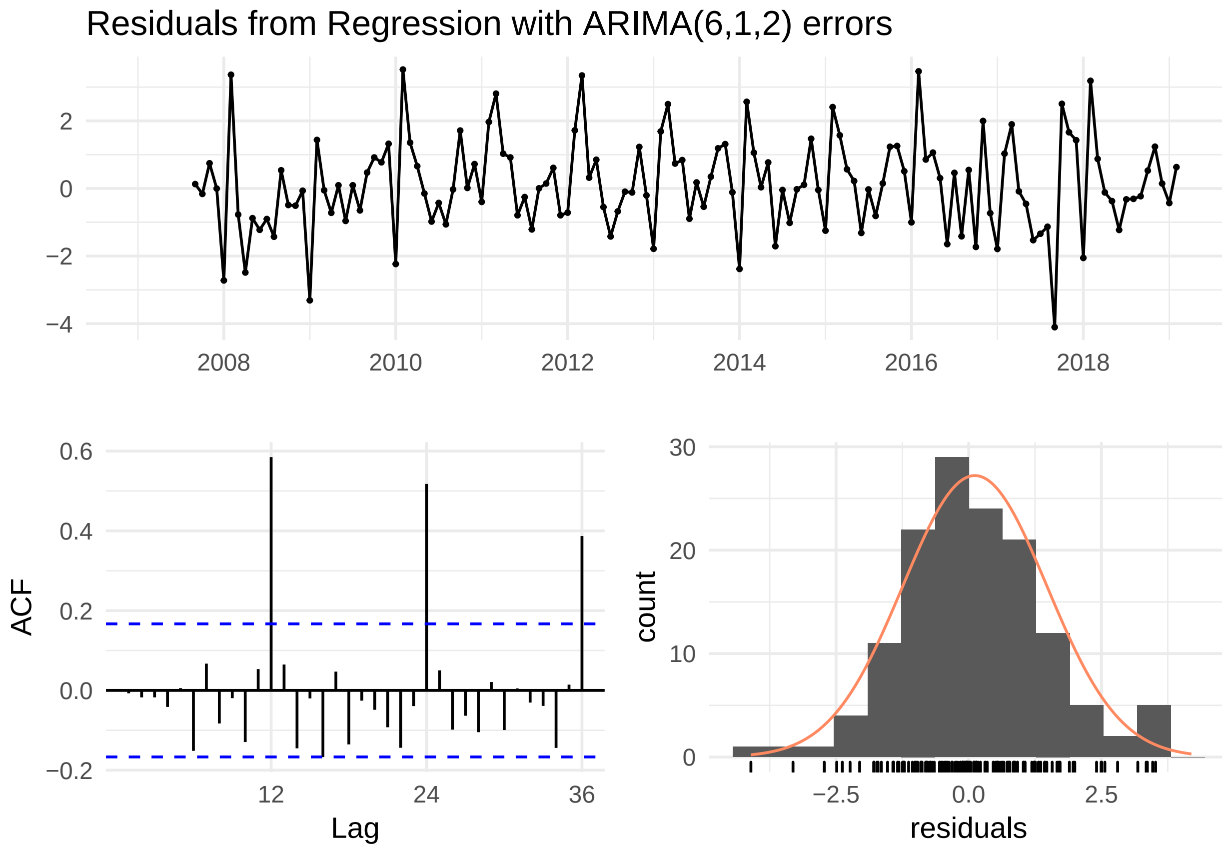
```
## 59 . L11_All_Employees
## 60 . L10_All_Employees
## 61 . L9_All_Employees
## 62 . L8_All_Employees
## 63 . L7_All_Employees
## 64 . L6_All_Employees
## 65 . L5_All_Employees
## 66 . L4_All_Employees
## 67 . L3_All_Employees
## 68 . L2_All_Employees
## 69 . L1_All_Employees
## 70 . L12_Retail_Employees
## 71 . L11_Retail_Employees
## 72 . L10_Retail_Employees
## 73 . L9_Retail_Employees
## 74 . L8_Retail_Employees
## 75 . L7_Retail_Employees
## 76 . L6_Retail_Employees
## 77 . L5_Retail_Employees
## 78 . L4_Retail_Employees
## 79 . L3_Retail_Employees
## 80 . L2_Retail_Employees
## 81 . L1_Retail_Employees
## 82 . L12_Local_Gov_Employees
## 83 . L11_Local_Gov_Employees
## 84 . L10_Local_Gov_Employees
## 85 . L9_Local_Gov_Employees
## 86 . L8_Local_Gov_Employees
## 87 . L7_Local_Gov_Employees
## 88 . L6_Local_Gov_Employees
## 89 . L5_Local_Gov_Employees
## 90 . L4_Local_Gov_Employees
## 91 . L3_Local_Gov_Employees
## 92 . L2_Local_Gov_Employees
## 93 . L1_Local_Gov_Employees
## 94 . L12_Healthcare_Employees
## 95 . L11_Healthcare_Employees
## 96 . L10_Healthcare_Employees
## 97 . L9_Healthcare_Employees
## 98 . L8_Healthcare_Employees
## 99 . L7_Healthcare_Employees
## 100 . L6_Healthcare_Employees
## 101 . L5_Healthcare_Employees
## 102 . L4_Healthcare_Employees
## 103 . L3_Healthcare_Employees
## 104 . L2_Healthcare_Employees
## 105 . L1_Healthcare_Employees
## 106 . L12_Federal_Employees
## 107 . L11_Federal_Employees
## 108 . L10_Federal_Employees
## 109 . L9_Federal_Employees
## 110 . L8_Federal_Employees
## 111 . L7_Federal_Employees
## 112 . L6_Federal_Employees
```

```
## 113 . L5_Federal_Employees
## 114 . L4_Federal_Employees
## 115 . L3_Federal_Employees
## 116 . L2_Federal_Employees
## 117 . L1_Federal_Employees
## 118 . Federal_Employees
## 119 . Healthcare_Employees
## 120 . Local_Gov_Employees
## 121 . Retail_Employees
## 122 . All_Employees
## 123 . Average_Hourly_Earnings
## 124 . Average_Weekly_Earnings
## 125 . Total_Weekly_Earnings
##
##
## Variables Entered:
##
## - L1_Average_Weekly_Hours
## - D1_Average_Weekly_Hours
## - L2_Local_Gov_Employees
## - D1_Total_Weekly_Earnings
## - D1_Average_Weekly_Earnings
##
## No more variables to be added.
```

```
##
##                                   Selection Summary
## -------------------------------------------------------------------------------------
## Variable                          AIC        Sum Sq      RSS       R-Sq      Adj. R-Sq
## -------------------------------------------------------------------------------------
## L1_Average_Weekly_Hours         227.678    141.021    39.161    0.78266      0.78114
## D1_Average_Weekly_Hours       -9101.524    180.182     0.000    1.00000      1.00000
## L2_Local_Gov_Employees        -9379.775    177.297     0.000    1.00000      1.00000
## D1_Total_Weekly_Earnings      -9384.215    177.297     0.000    1.00000      1.00000
## D1_Average_Weekly_Earnings    -9384.887    177.297     0.000    1.00000      1.00000
## -------------------------------------------------------------------------------------
```

For total employment I will be using the following model:

- Lags 1,3,6,8 of Local government employment, federal government employment, retail employment, healthcare employment
- 6 lags of Total Employment
- Lags 1,3,8 of Average Weekly Hours
- Differencing once
- 2 Moving average terms




```
##
##  Ljung-Box test
##
## data:  Residuals from Regression with ARIMA(6,1,2) errors
## Q* = 123.63, df = 15, p-value < 2.2e-16
##
## Model df: 9.   Total lags used: 24
```

```
## # A tibble: 3 x 6
##   `Point Forecast` `Lo 80` `Hi 80` `Lo 95` `Hi 95` Date
##              <dbl>   <dbl>   <dbl>   <dbl>   <dbl> <chr>
## 1             186.    182.    190.    179.    193. March
## 2             186.    182.    191.    179.    194. April
## 3             187.    182.    192.    179.    194. May
```

This model is incomplete. There is still serial correlation and after many iterations of the model and even after running exhaustive automated model selection we were not able to find a complete model.

---

## Prophesying

In this section I'll be using techniques not covered in class. The team at Facebook published a paper and a complementary package on time series. The *Prophet* method works by using additive regression with four main components:

1. A piece-wise linear or logistic growth curve trend
2. A yearly seasonal component modeled using Fourier series
3. A weekly seasonal component using dummy variables ( not applicable )
4. A user provided list of important holidays

### Average Weekly Hours

```
## Parsed with column specification:
## cols(
##   DATE = col_date(format = ""),
##   Federal_Employees = col_double(),
##   Healthcare_Employees = col_double(),
##   Local_Gov_Employees = col_double(),
##   Retail_Employees = col_double(),
##   All_Employees = col_double(),
##   Average_Weekly_Hours = col_double(),
##   Average_Hourly_Earnings = col_double(),
##   Average_Weekly_Earnings = col_double(),
##   Total_Weekly_Earnings = col_double()
## )
```

```
## Disabling weekly seasonality. Run prophet with weekly.seasonality=TRUE to override this.
```

```
## Disabling daily seasonality. Run prophet with daily.seasonality=TRUE to override this.
```

```
##             ds    trend
## 147 2019-03-01 32.49304
## 148 2019-04-01 32.46231
## 149 2019-05-01 32.43258
```

### Average Hourly Earnings

```
## Disabling weekly seasonality. Run prophet with weekly.seasonality=TRUE to override this.
```

```
## Disabling daily seasonality. Run prophet with daily.seasonality=TRUE to override this.
```

```
##             ds    trend
## 147 2019-03-01 22.19575
## 148 2019-04-01 22.30230
## 149 2019-05-01 22.40541
```

### Total Employment

```
## Disabling weekly seasonality. Run prophet with weekly.seasonality=TRUE to override this.
```

```
## Disabling daily seasonality. Run prophet with daily.seasonality=TRUE to override this.
```

```
##             ds    trend
## 147 2019-03-01 182.9895
## 148 2019-04-01 183.2797
## 149 2019-05-01 183.5605
```

# Results

### Average Weekly Hours

| | ds | yhat | yhat_lower | yhat_upper |
|---|---|---|---|---|
| **147** | 2019-03-01 | 32.97 | 32.38 | 33.58 |
| **148** | 2019-04-01 | 32.88 | 32.3 | 33.45 |
| **149** | 2019-05-01 | 32.44 | 31.85 | 32.98 |

### Average Hourly Earnings

| | ds | yhat | yhat_lower | yhat_upper |
|---|---|---|---|---|
| **147** | 2019-03-01 | 22.53 | 21.88 | 23.12 |
| **148** | 2019-04-01 | 22.42 | 21.73 | 23.11 |
| **149** | 2019-05-01 | 22.44 | 21.77 | 23.11 |

### Total Employment

| | ds | yhat | yhat_lower | yhat_upper |
|---|---|---|---|---|
| **147** | 2019-03-01 | 184.7 | 183.8 | 185.6 |
| **148** | 2019-04-01 | 184.9 | 184 | 185.7 |
| **149** | 2019-05-01 | 184.5 | 183.6 | 185.4 |

# Finalized Results

Table 8: Table continues below

| Date | Total_Employment | Average_Hourly_Earnings | Average_Weekly_Hours |
|---|---|---|---|
| 2019-01-01 | 179.9 | 21.94 | 32.6 |
| 2019-02-01 | 183.4 | 22.2 | 32.67 |
| 2019-03-01 | 184.7 | 22.53 | 32.97 |
| 2019-04-01 | 184.9 | 22.42 | 32.88 |
| 2019-05-01 | 184.5 | 22.44 | 32.44 |

| Average_Weely_Earnings | Total_Weekly_Earnings |
|---|---|
| 715.1 | 128642 |
| 725.4 | 133029 |
| 742.9 | 137190 |
| 737.2 | 136291 |
| 728 | 134310 |

For the upper and lower bounds refer to the tables above.

# Conclusions

The final forecast is strong in the sense that I have confidence in the numbers. Since it's a pre-packaged and highly complex algorithm that generates the forecast I don't the the mathematical prowess to explain it's inner workings. Some of the latter models I can explain. They all have limitations and within the upper and lower bounds generated by the forecast we can safely take action on those predictions. All the models I made except the dynamic model for Total Employment were complete and predicted accurately on test data. I can't confirm this for a rolling window root mean squared error since I didn't have the proper tooling to complete the calculation. I would like to create a function of my own to create such a metric but due to time constraints was unable to.

With that being said I would like to reiterate by saying that I believe the final forecast, and those before it are complete. If the `prophet` package is not applicable for the evaluation for this work then please refer back to the previous models and use those for the evaluation. Although date/month labels are not provided, the forecasts are in chronological order.

## Using R

I hurt my hand from the amount of times I banged my fist against my desk trying to run models and figure out time series in R. However, I am glad I forced myself to do so. Not only have I learnt skills that I can immediately apply to my current and future work, but I also understand a little more about R series analysis. I've learnt about Box-Cox transformations, naive, seasonal naive, Ljung-Box tests and many more techniques that I aim to use in the future. Regardless of the final evaluation of this project I feel like I have achieved a lot.