PROJECT II

ROWAN CURRY

INTRODUCTION

The purpose of this report is to analyze a dataset using as many different methods learned in this course as possible. The intention of this project is to explore a time series object in an attempt to practice new technical skills while making meaningful inferences about the world.

In the following pages, we start with the "Material and Methods" section, where the report describes the data qualitatively and quantitatively through exploratory statistical analysis. This section also addresses the stationarity of the data and the sample ACF and PACF that will be referenced in the final section.

The final section can be located under "Results", and it is here where the ARMA models will be built and analyzed. It is also here where we'll draw conclusions about which models are best, and make any inferences that we can about the data as a result of our analysis.

MATERIAL AND METHODS

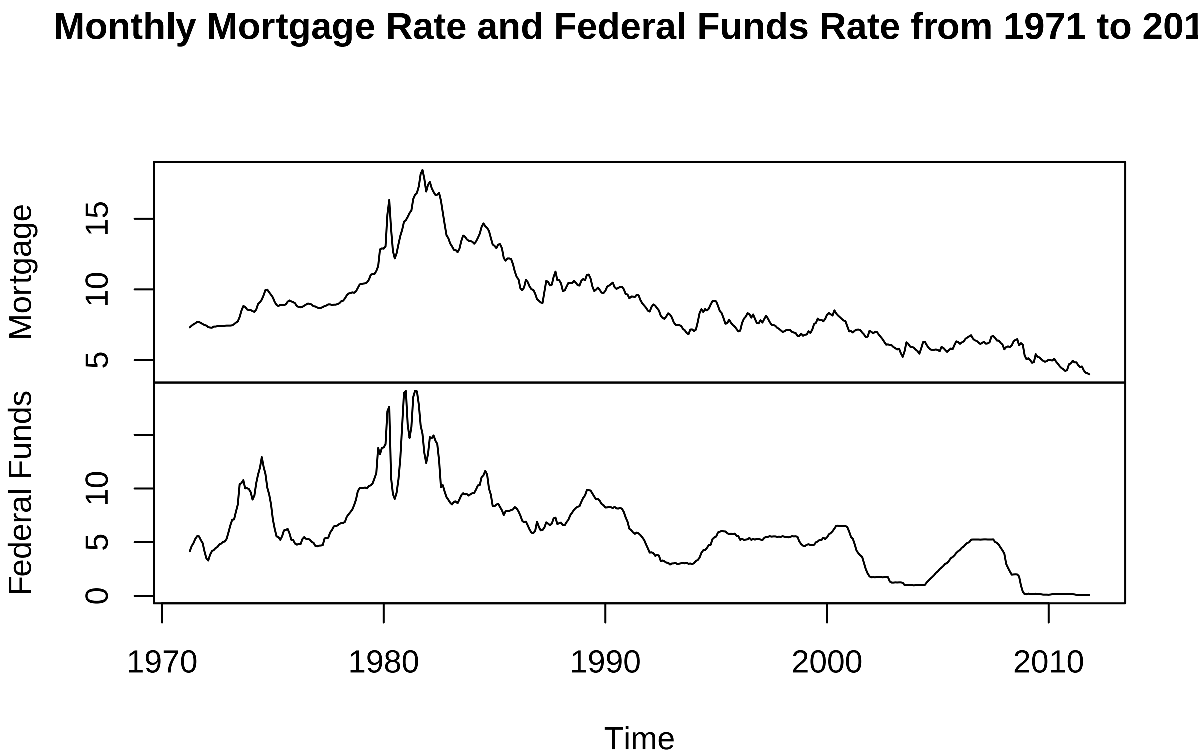
Data Description

The data used in this report is from the US monthly 30-year conventional mortgage rates from April 1971 to November 2011. The data consists of five columns– *year, month, day, morg, ffr.* The column *"morg"* stands for monthly mortgage rates, while the column *"ffr"* contains the monthly federal funds rate.

This data is a time series because it's a sequence of numerical data points in successive, equally spaced order. Each variable is recorded once per month during a year long period, so each combination of year plus variable results in two time series– one for mortgage, and one for the monthly federal funds rate.

Exploratory Data Analysis

The plot of the two time series is displayed below.



We can see from this graph that there is a dramatic spike in both monthly mortgage rates and federal funds rates from approximately 1979 - 1984. This means that these years will most likely be significant when it comes to our analysis. There is also a smaller spike in rates around the year 2008, which means that we can hypothesize that a decrease in mortgage and federal funds antecedes a recession, and an increase in these rates is characteristic of these economic events. The similarity of these two plots also suggests some kind of correlation between the two datasets.

We can explore our data further by investigating the seasonality of the data and fitting a linear regression model to the consequent time series.

The coefficients from the seasonal model of monthly mortgage rates are shown below.

```
##           Coefficient
## January   -0.15356916
## February  -0.02170257
## March      0.06359486
## April      0.06115649
## May        0.09371002
## June       0.08626355
## July       0.08759756
## August     0.11527304
## September  0.11319242
## October    0.10745326
## November   0.06025069
## December   0.03820257
```

We can see from the coefficients above that mortgage rates tend to decrease in January and February, and then generally increase until they reach their peaks in August and September, and then begin to decrease again as winter approaches and we near January and February again.

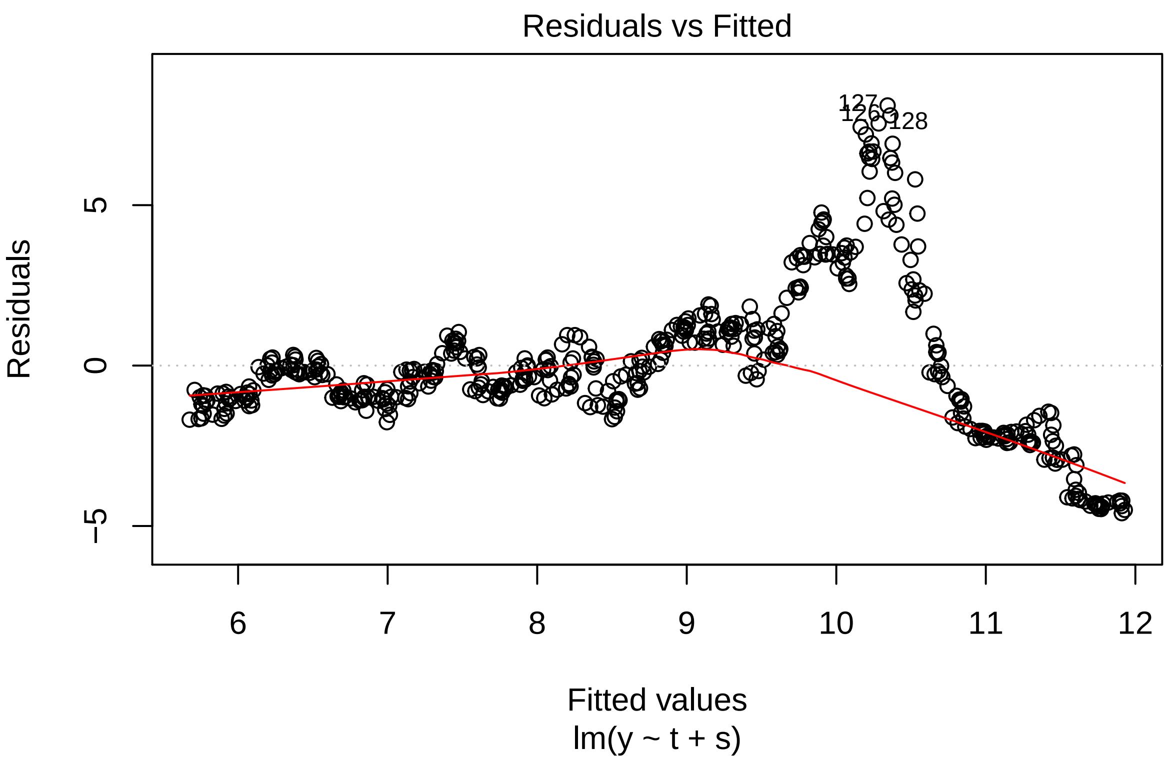
The coefficients from the seasonal model of federal funds rates are shown below.

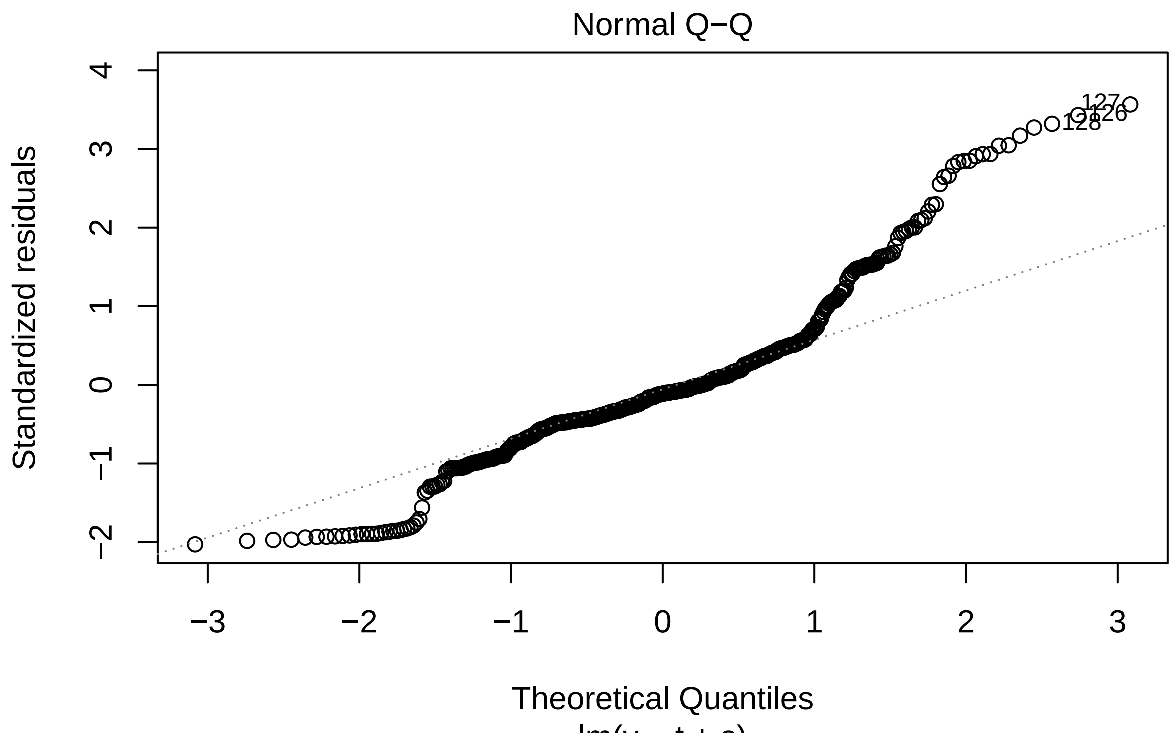
```
##            Coefficient
```

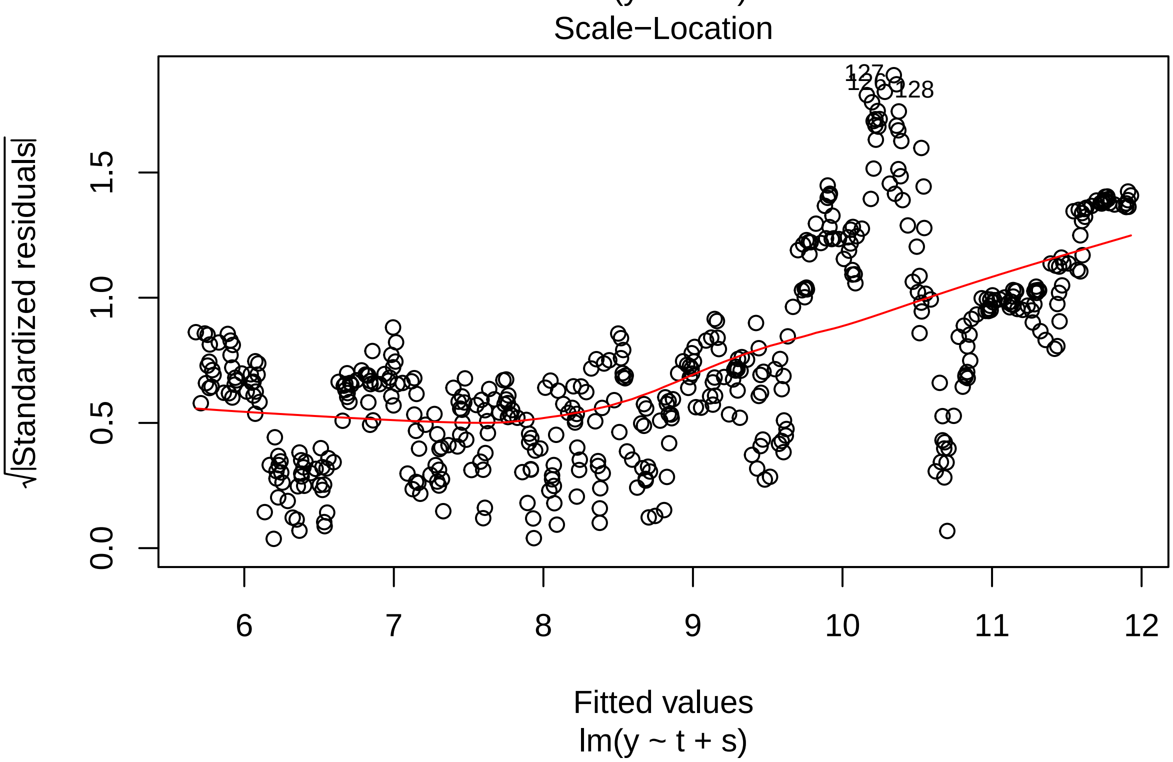
```
## January   -0.205939471
## February  -0.081588377
## March      0.022073245
## April     -0.043350721
## May       -0.038140318
## June       0.048533499
## July       0.118378049
## August     0.089686013
## September  0.110262269
## October    0.057667794
## November  -0.001268144
## December   0.044838377
```

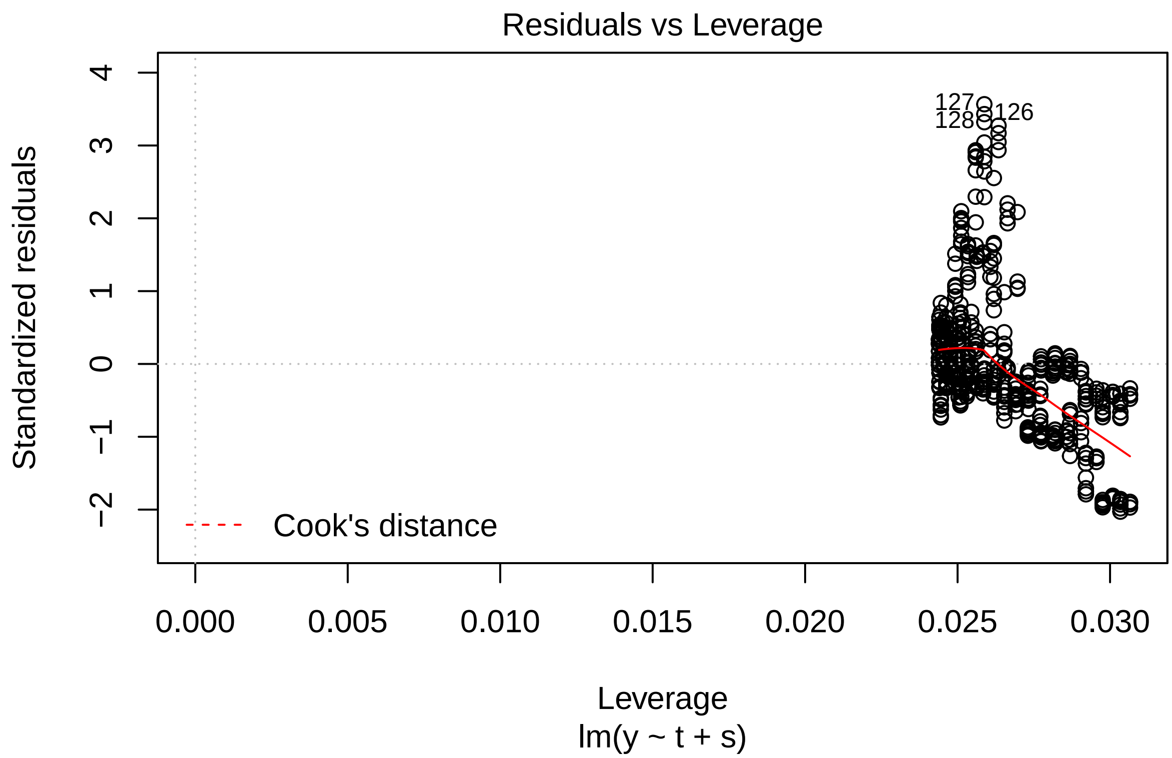
We can see from the coefficients above that federal funds rates tend to decrease in January and February, similar to the monthly mortgage rates, and then increase in March– however, they tend to decrease again in April and May, before climbing back up in June and onward. They begin to decrease again in November, but then quickly go back up in December. This seems to be less of a straightforward trend when compared to the seasonality coefficients for mortgage rates.

The plots for the seasonal regression model for monthly mortgage rates are displayed below.



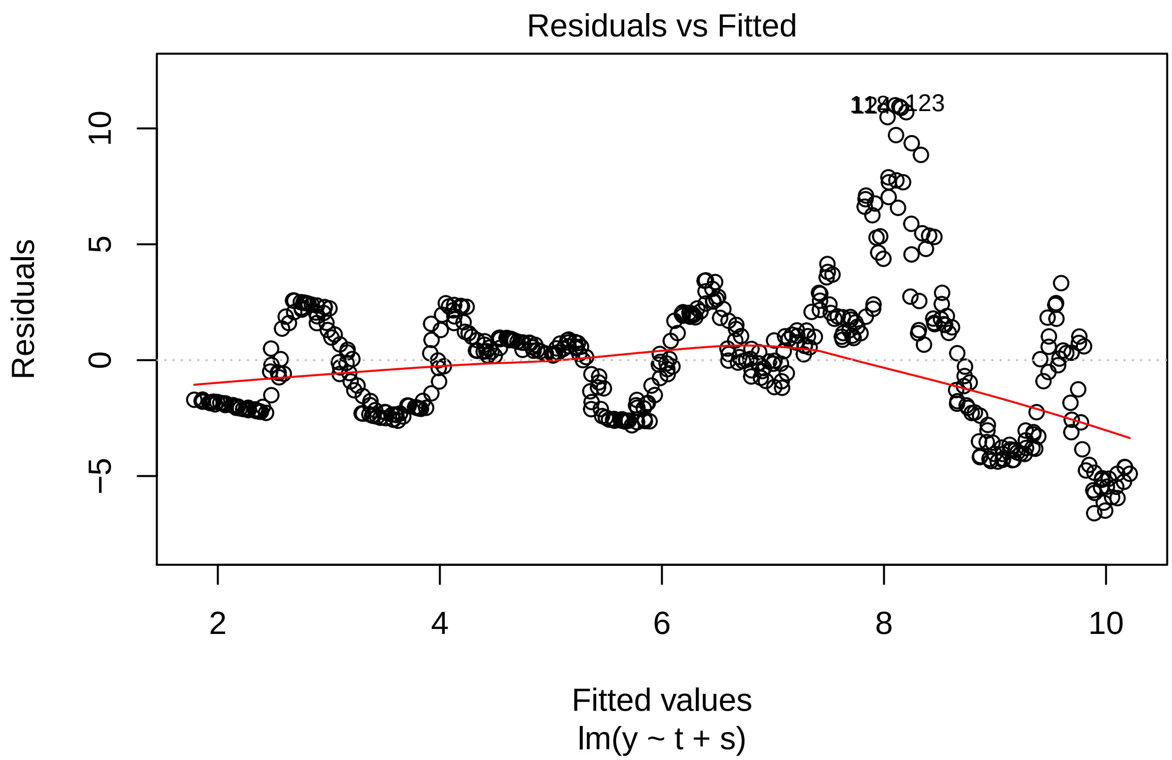

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(y ~ t + s)
127
126
128

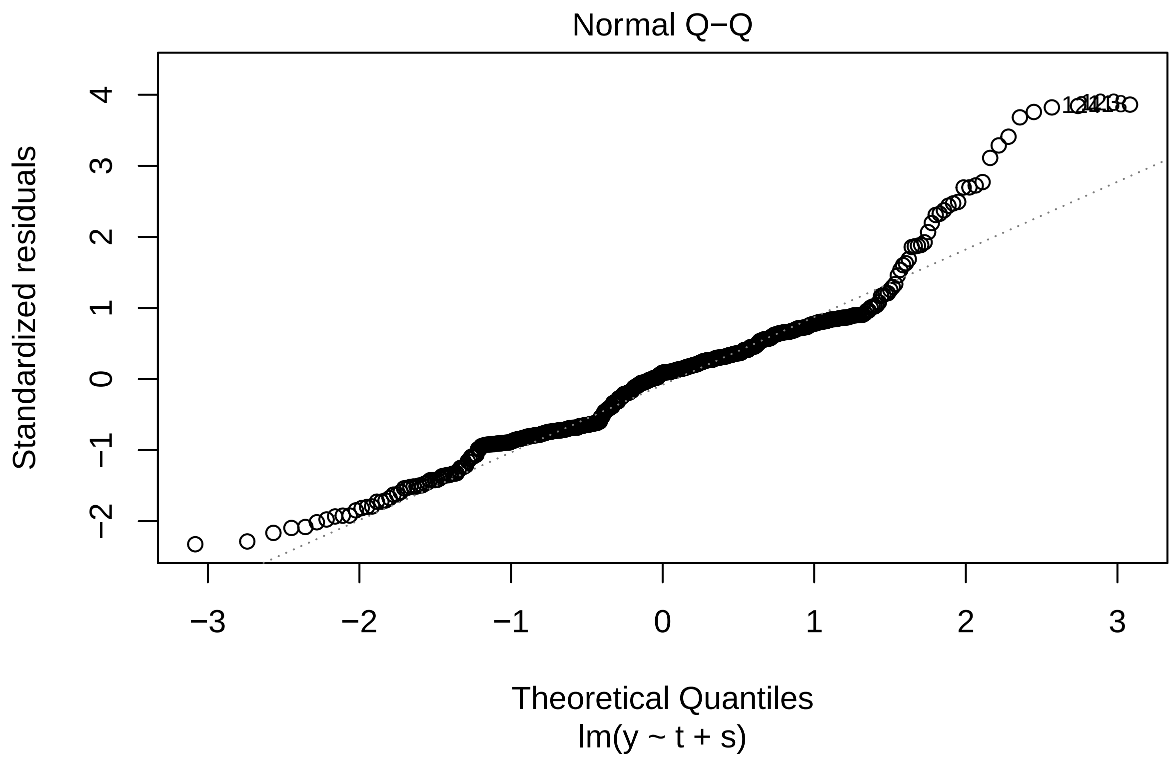

Scale−Location
√|Standardized residuals|
Fitted values
lm(y ~ t + s)
127
126
128

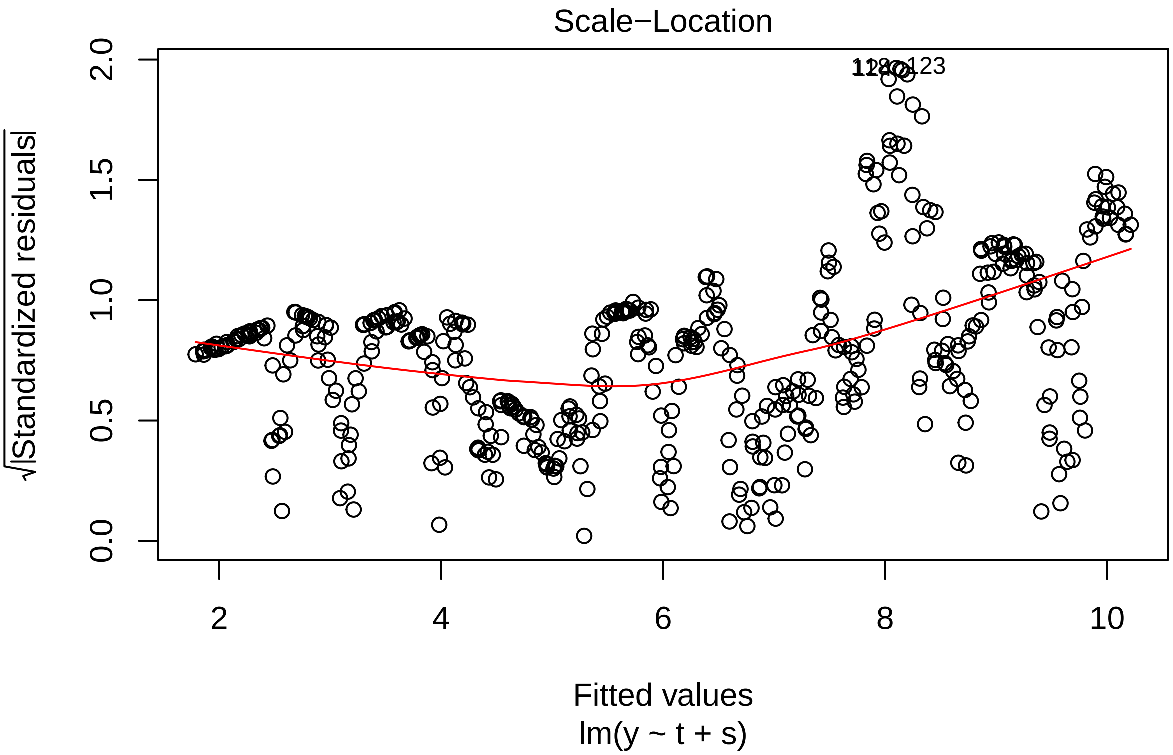


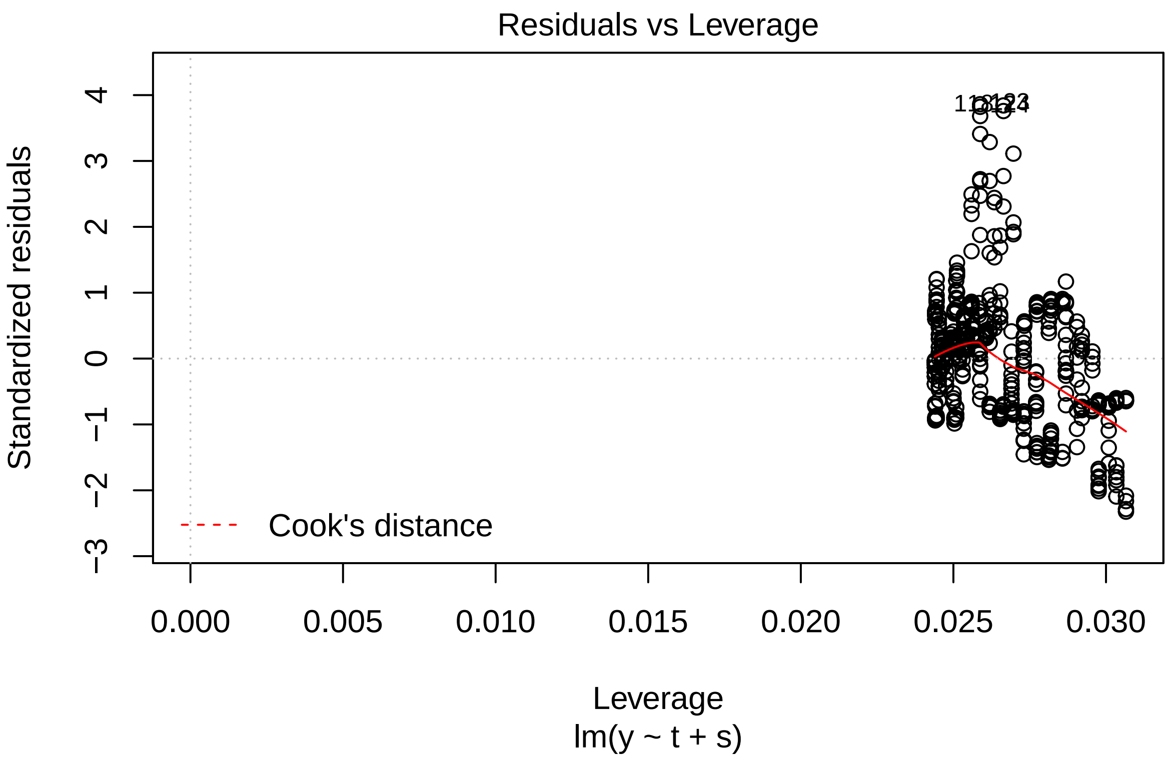
We can see from the Residuals vs. Fitted plot that the data points are in no way evenly distributed around the 0 axis, which implies that there is no linear trend in this data. When we observe the Normal Q-Q plot, we can see that while some of the data follows a normal distribution, it departs from normality in significant ways.

The plots for the seasonal regression model for monthly federal funds rates are displayed below.




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(y ~ t + s)
−3
−2
−1
0
1
2
3
4


Scale−Location
√|Standardized residuals|
Fitted values
lm(y ~ t + s)
123
0.0
0.5
1.0
1.5
2.0
2
4
6
8
10

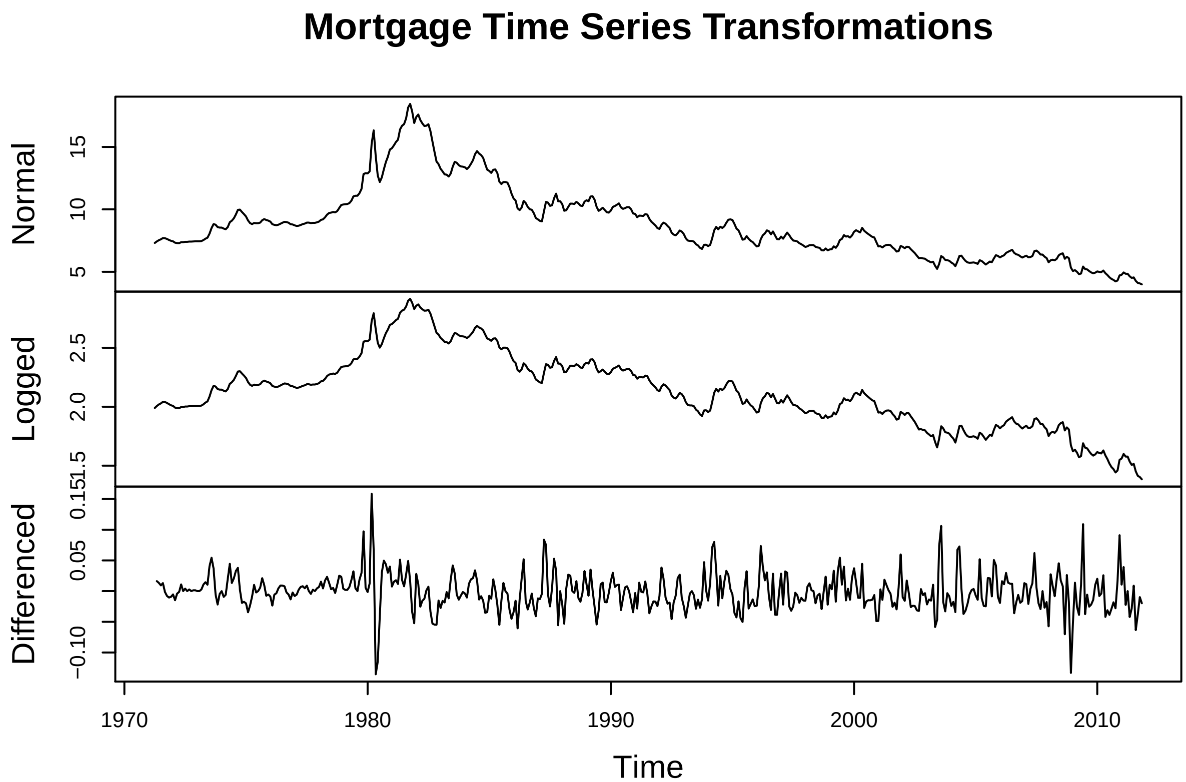


We can observe very similar results among the above graphs as we did for the results of applying seasonal regression to the monthly mortgage rates. While the residuals vs. fitted plot suggests that this regression model might be slightly closer to a linear trend than the previous model, the distribution of residuals still suggests a departure from linearity. The Normal QQ plot here also leads to similar conclusions as the Normal QQ plot for the previous model.

Stationarity

By referencing the plot of the Monthly Mortgage Rate Time Series given at the very beginning of the Materials & Method section, we can clearly observe that the monthly mortgage rate series is not stationary. This is because from approximately 1983 to 2011 we can observe a linear trend. We can also observe that variance is not constant– this can be seen by the varying degrees by which the graph dips and spikes.

Because this time series is not stationary, we will test different transformations in order to determine which one can help make the time series stationary. The resulting plot of these transformation tests is shown below.

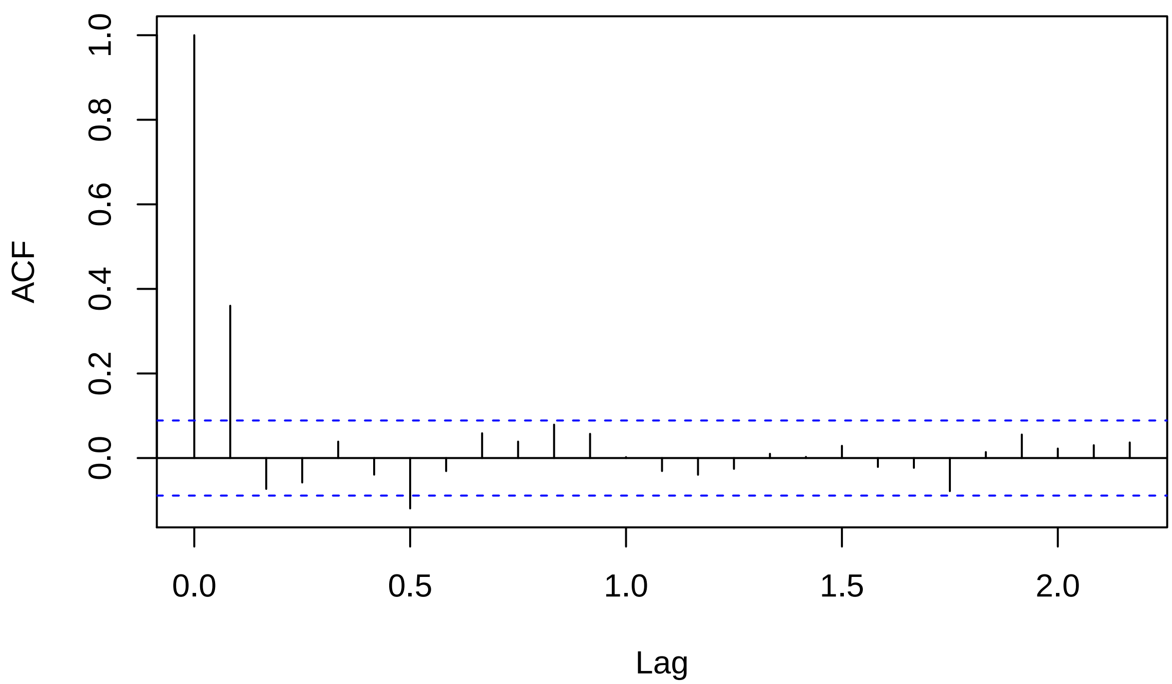


We can see from the plot above that applying a differenced log transformation results in a stationary series. This transformation can be observed in the lowermost transformation, where we can see that the data has no observable trend, and the variance remains relatively constant throughout. We can see in the next plot that applying a log transformation does little to result in a stationary series. The plot of the normal time series was provided above the two transformations for reference.

Sample ACF & PACF

The sample ACF for the monthly mortgage rates is provided below.

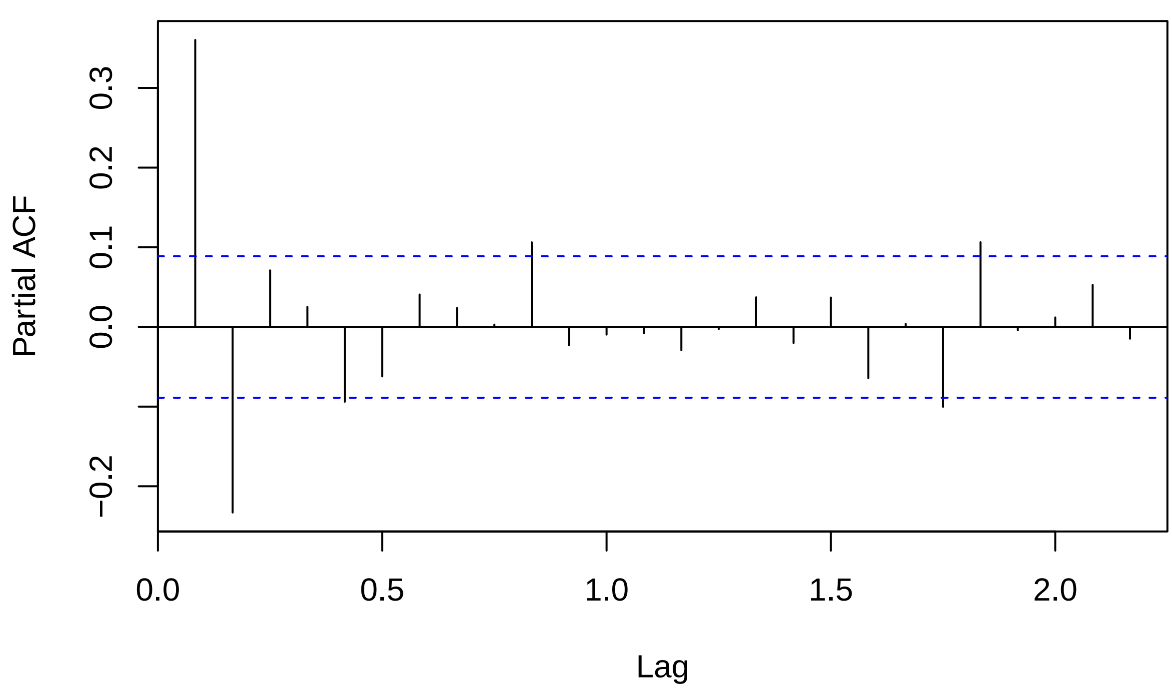
Sample ACF for Monthly Mortgage Rates



We can observe from the plot of the Sample ACF above that the ACF cuts off before the lag reaches 1. This is characteristic of several different kinds of models, so we will have to observe the Sample PACF as well in order to make an educated guess about what kind of model is suggested by these functions.

The sample PACF for the monthly mortgage rates is provided below.

Sample PACF for Monthly Mortgage Rates



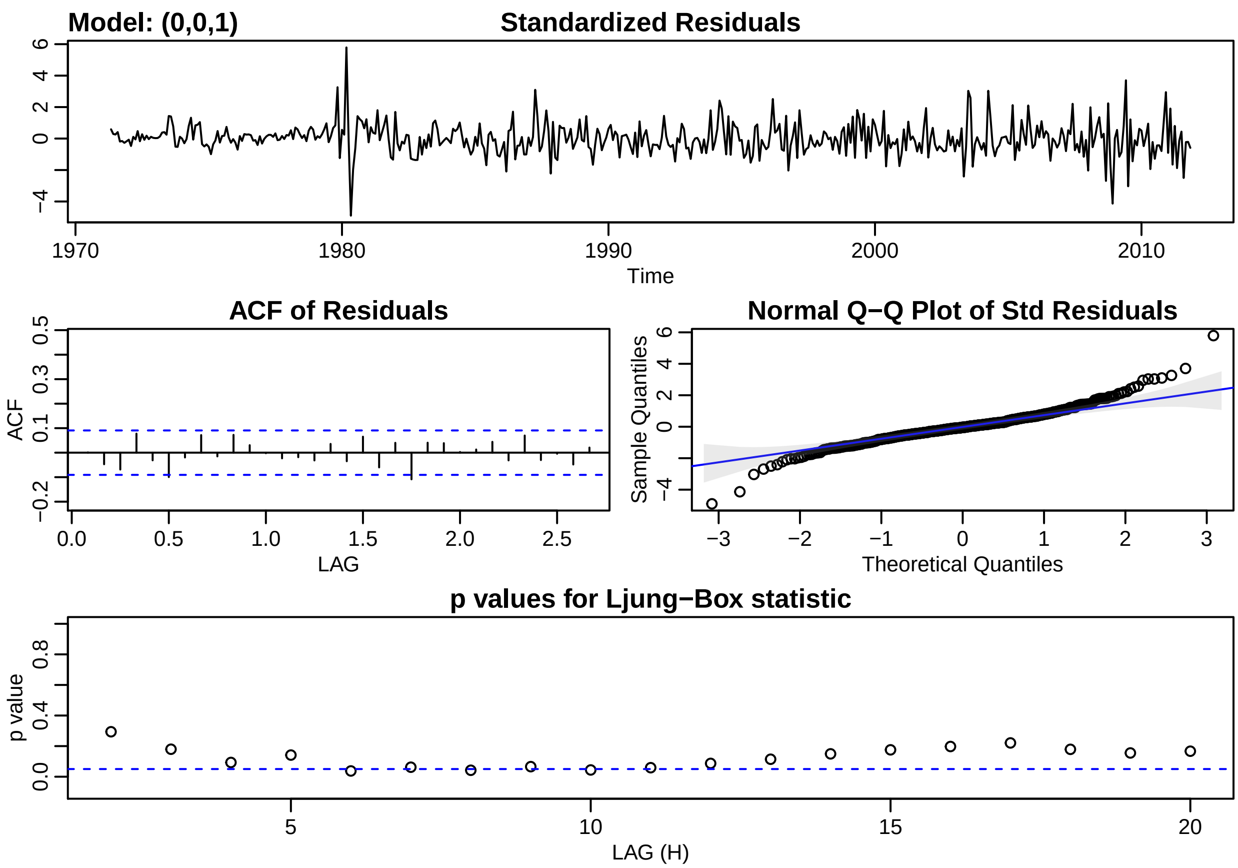
We can see from the plot above that the Sample PACF cuts off around lag 0.25. This would suggest that the data fits an AR(p) model, where p represents lag and p = 0, since lag cuts off before 1.

RESULTS

Building an ARMA Model

We saw in the previous section that the Sample ACF and Sample PACF suggested a model of AR(p), where p = 0. We also know that in AR(p) models, q = 0. This means all we have left to figure out when it comes to choosing an ARMA model is the value of d, which is represented by the number of times you had to difference the time series in order to remove the trend. In the Stationarity section of this report, we saw that differencing the data only once resulted in a stationary time series. Therefore we have a value of d = 1, which gives an ARMA model with p = 0, d = 1, and q = 0.

To check whether this is an appropriate model, we can calculate the autocorrelation of the ARMA model ARMA(0,0,1) by calculating the fitted model. The results are provided below.



We can see here that the autocorrelation is zero, which means that our residuals look white (they mimic a white noise process). Additionally, the Normal QQ plot suggests that our results are approximately Normal. This means that the ARMA(0,0,1) model is a good fit, and we can feel confident using it to examine our results moving forward.

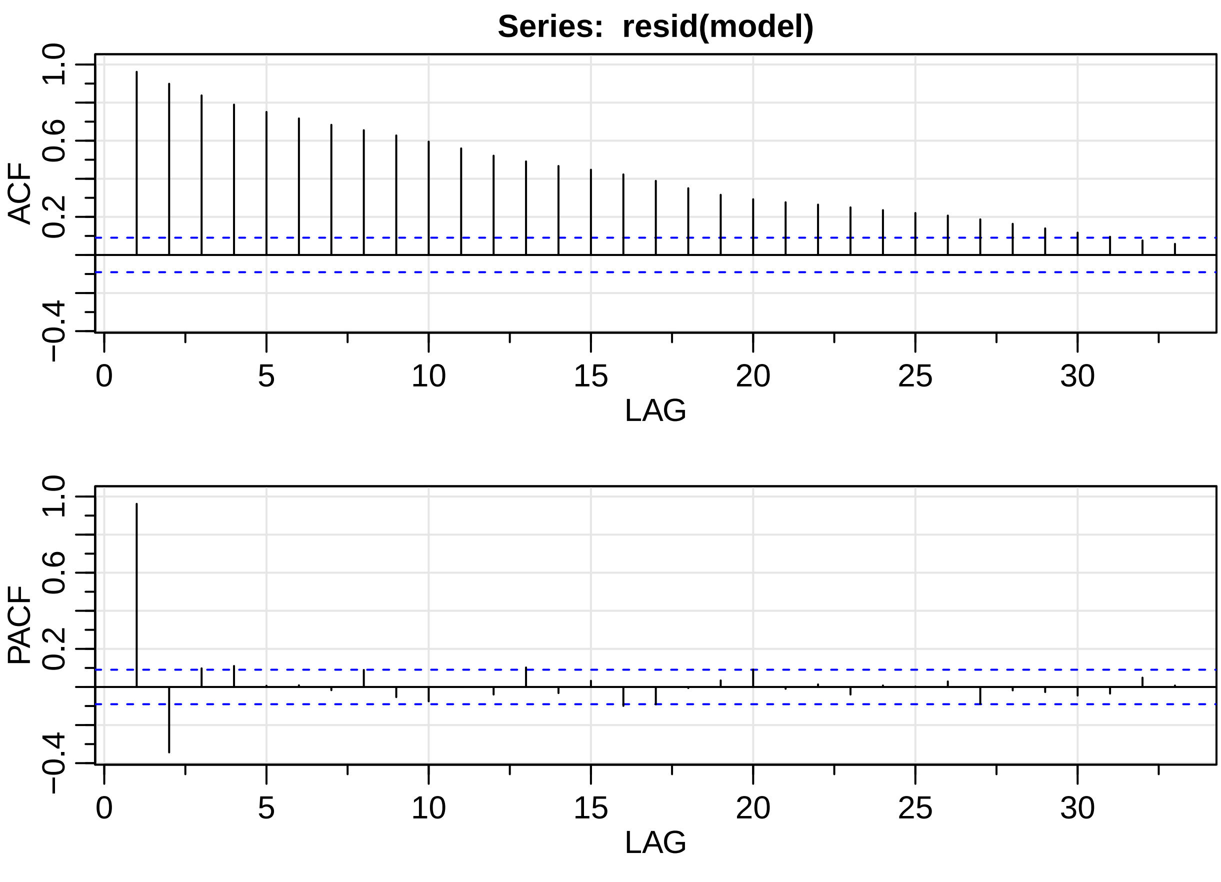
Building an ARMA Model for Mortgage Rates and lag-1 Federal Funds Rates

First, we calculate the lag-1 Federal Funds Rates. Then, we use this new data as an explanatory variable in a new time series model for the mortgage rate. The summary of this model is provided below.

```
##
```

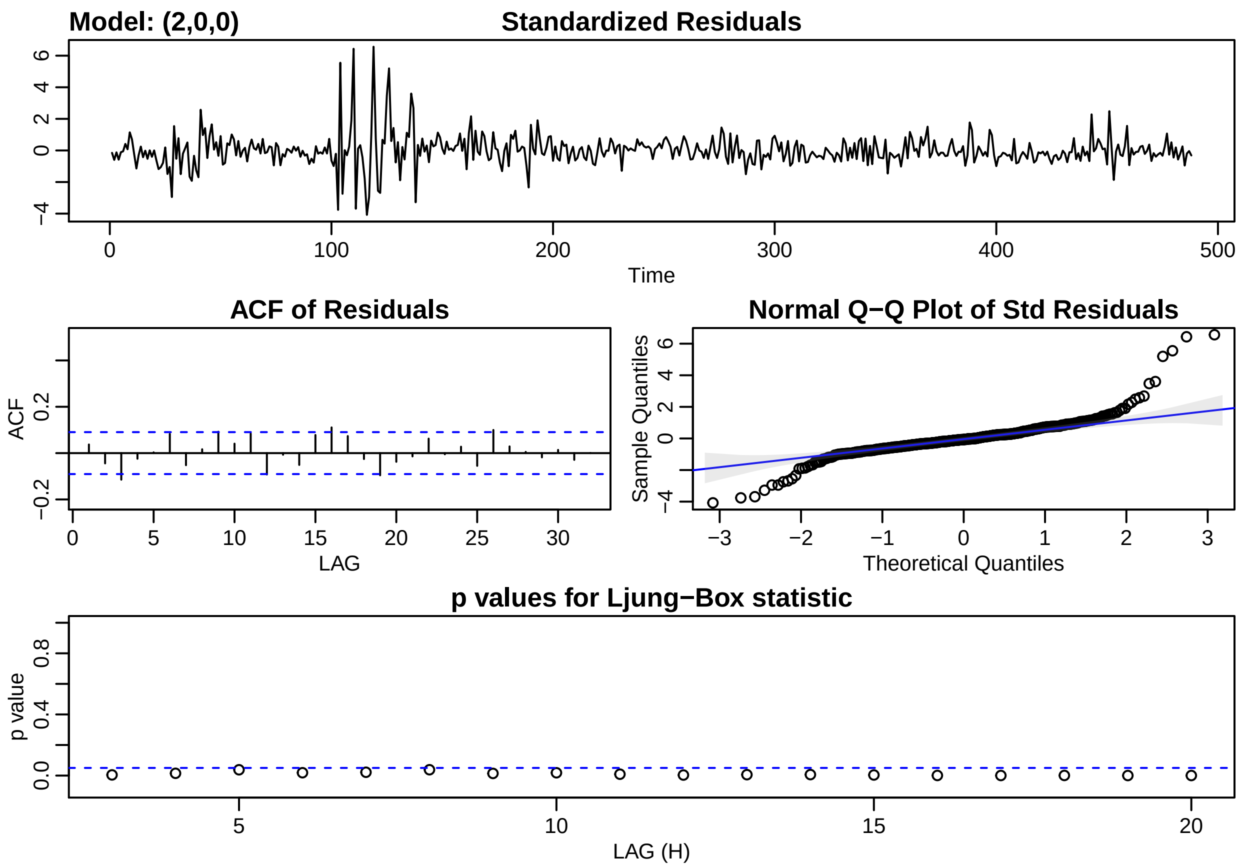
```
## Call:
## lm(formula = morg.ts ~ lag.ffr.ts, data = morg.ffr.lag)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2872 -0.8823 -0.0294  0.7728  4.6306
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.67167    0.11505   40.60   <2e-16 ***
## lag.ffr.ts   0.68851    0.01629   42.27   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.344 on 486 degrees of freedom
## Multiple R-squared:  0.7862, Adjusted R-squared:  0.7857
## F-statistic:  1787 on 1 and 486 DF,  p-value: < 2.2e-16
```

Now that we have our fitted model, we can plot the Sample ACF and PACF with the intention of building an appropriate time series model for this new form of data.



We can see very clearly from the plot of the ACF, which tails off, and the PACF, which cuts off after approximately lag p = 2, that these plots suggest an AR(2) model.

This means we have a value of p = 2, and q = 0. The use of lag - 1 without any transformations when fitting the model suggests a value of d = 0. Therefore we can go on to test an ARMA model as we did in the previous section. In this case, it will be the ARMA(2,0,0) model, and the fitted results are provided below.



We can see here that once again, the autocorrelation is zero, and our residuals look white. Additionally, the Normal QQ plot suggests that our results are approximately Normal. This means that the ARMA(2,0,0) model is a good fit.

This model suggests a strong correlation between mortgage rates and the federal funds rate.

APPENDIX

```
#########################################################################################
# LOADING THE DATA
data <- read.delim("~/Desktop/mortgage.txt", sep = " ")
# HELPFUL PACKAGES
library(astsa)
library(zoo)
#########################################################################################
# EXPLORATORY ANALYSIS
# converting data to time series
morg.ts <- ts(data$morg, frequency = 12, start = c(1971,4))
ffr.ts <- ts(data$ffr, frequency = 12, start = c(1971,4))
# combining the time series
combined <- cbind(morg.ts, ffr.ts)
# renaming variables so that the plot looks nicer
colnames(combined) <- c("Mortgage", "Federal Funds")
# plotting the time series in the same graph
plot(combined, main = "Monthly Mortgage Rate and Federal Funds Rate from 1971 to 2011")
# seasonality, linear regression
```

```r
# functions to help with data analysis
df_fxn <- function(x){
  df = data.frame( # function to convert time series
    y = x,         # into a data frame
    t = time(x),
    s = as.factor(cycle(x))
  )
}
seasonality.lm <- function(x) {
  df = df_fxn(x) # function to apply seasonal regression
  lm(y ~ t + s, data = df)
}
# applying seasonality.lm function
seasonal.morg <- seasonality.lm(morg.ts)
# table to make coefficients look nice in report
smorg.coeff <- matrix(c(seasonal.morg$coefficients[2:13]))
colnames(smorg.coeff) <- c("Coefficient")
rownames(smorg.coeff) <- c("January","February", "March", "April", "May", "June",
                      "July", "August", "September", "October", "November",
                      "December")
seasonal.ffr <- seasonality.lm(ffr.ts)
# table to make coefficients look nice in report
sffr.coeff <- matrix(c(seasonal.ffr$coefficients[2:13]))
colnames(sffr.coeff) <- c("Coefficient")
rownames(sffr.coeff) <- c("January","February", "March", "April", "May", "June",
                      "July", "August", "September", "October", "November",
                      "December")
smorg.coeff
sffr.coeff
# plots
plot(seasonal.morg)
plot(seasonal.ffr)
########################################################################################
# STATIONARITY - monthly mortgage rate series only
# testing transformations
log.morg.ts <- log(morg.ts) # log transformation
diff.morg.ts <- diff(log(morg.ts)) # differenced log transformation
transformations <- cbind(morg.ts, log.morg.ts, diff.morg.ts)
colnames(transformations) <- c("Normal", "Logged", "Differenced")
plot(transformations, main = "Mortgage Time Series Transformations")
########################################################################################
# SAMPLE ACF & PACF - monthly mortgage rate series only
# plots, using the transformed data
acf(diff.morg.ts, na.action = na.omit, main = "Sample ACF for Monthly Mortgage Rates")
pacf(diff.morg.ts, na.action = na.omit, main = "Sample PACF for Monthly Mortgage Rates")
########################################################################################
# ARMA MODEL - monthly mortgage rates only, model checking
astsa::sarima(diff.morg.ts, 0, 0, 1)
########################################################################################
# TIME SERIES MODEL - FEDERAL FUNDS LAG -1
lag.ffr.ts <- lag(ffr.ts, 1) # did 1 bc -1 means move forward 1
morg.ffr.lag <- as.zoo(intersect(morg.ts, lag.ffr.ts))
# fit the model
```

```r
summary(model <- lm(morg.ts ~ lag.ffr.ts, data = morg.ffr.lag))
# plotting acf and pacf
acf2(resid(model))
astsa::sarima(resid(model), 2, 0, 0)
```