STA 137 Time Series Analysis
Project II

Yuehe Wen
914762463
12/2/2020

1 Introduction

The mortgage rate is a primary consideration for homebuyers looking to finance a new home purchase with a mortgage loan. One of the general accepted factors that affects the mortgage rate is the Federal Funds rate. As the mortgage rate is a typical time series dataset, this report has two aims: analyzing the characteristics of the mortgage data as well as examining the relationship between the mortgage rate and the Federal Funds rate.

This report will first analyze and describe the patterns of the data in the Material and Methods section. Then, with the data and methods, the next section, Results, will address the building and selecting process for the desired time series models to fit the mortgage data and examine whether the mortgage rate depends on the Federal Funds rate or not. The Results section will also interpretate the results of the analysis.

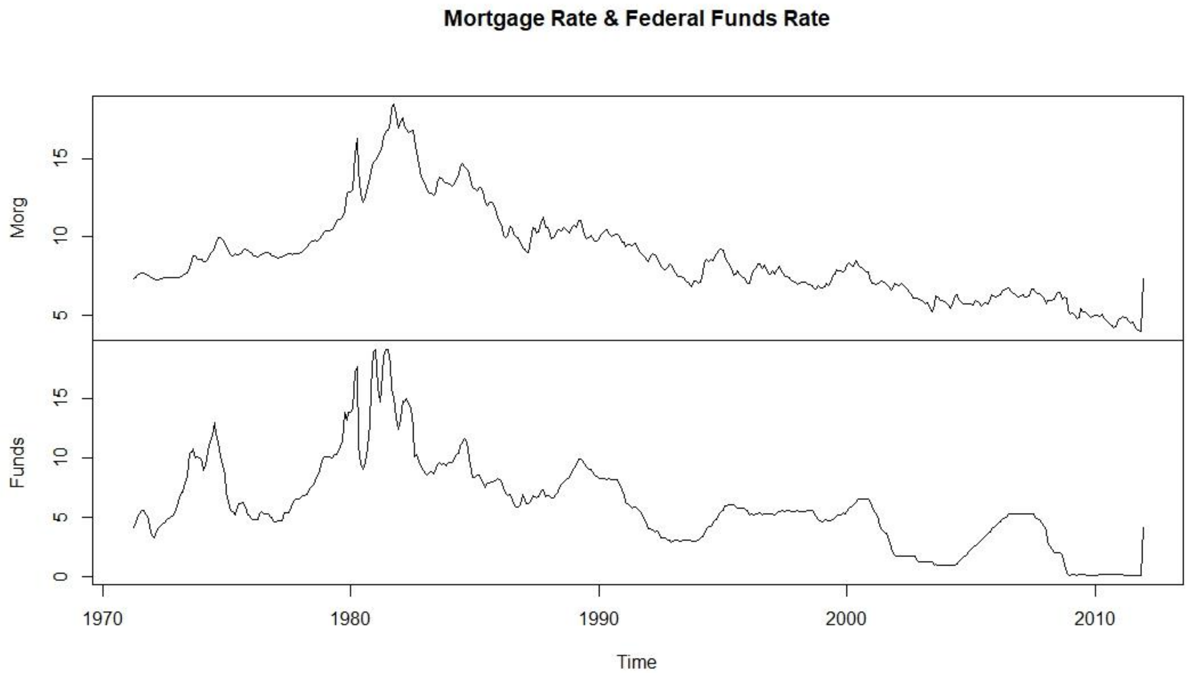
2 Material and Methods

This section will describe the patterns of the data and methods used for analysis.

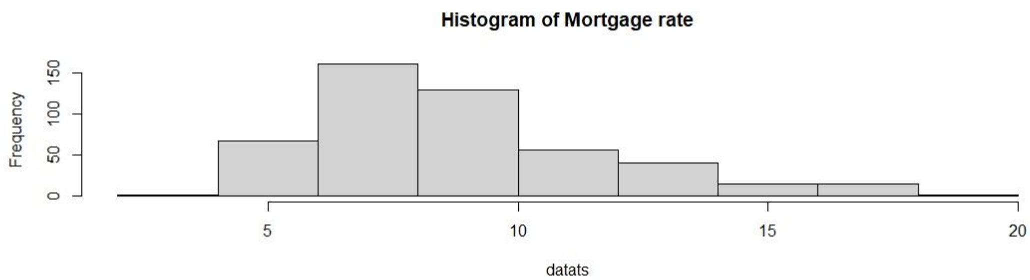
2.1 Material

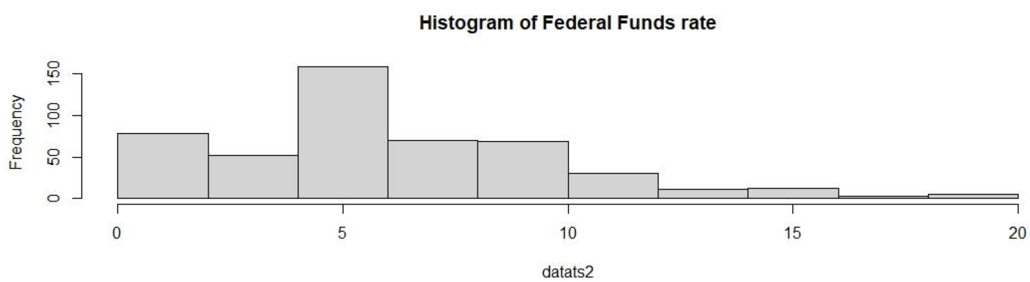
The dataset used is the US monthly 30-year conventional mortgage rates from April 1971 to November 2011. The source of the dataset is from the Board of Governors of the Federal Reserve System. The dataset consists of 5 columns (year, month, day, morg, ffr), where "morg" is the monthly mortgage rate and "ffr" is the monthly federal funds rate also collected by the Board of Governors of the Federal Reserve System.

The two rates are collected monthly, indicating that they have a sequence taken at successive equally spaced points in time. And the values are discrete. Therefore, the data is a time series data. The plot of the data is:



The histogram of the two rates is:





From the first plot, we can see that the two rates have same trend: the rates increased from 1971, both reached a highest peak around 1980 and decreased in the following years. Therefore, the series does not have a constant mean. There is no obvious seasonality in the mortgage rate, while in federal funds rate, the values go from high to low for two cycles in the period of 1990 – 2010.

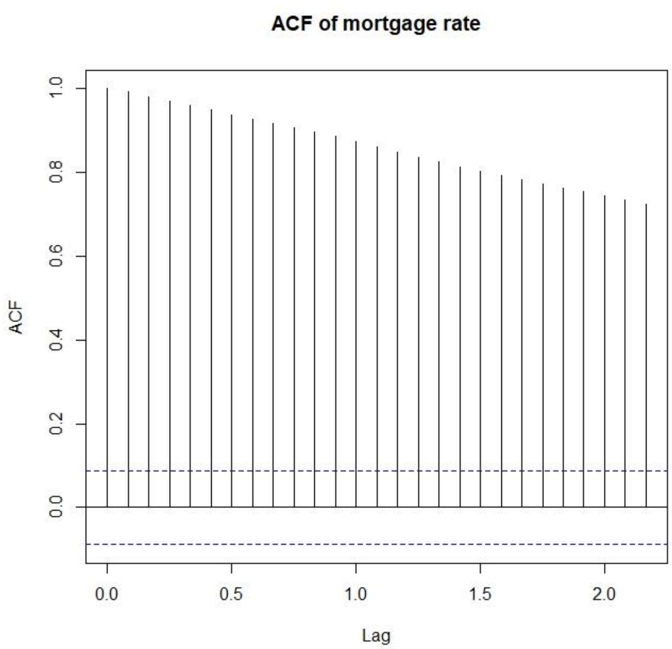
From the histograms, the frequencies of the values skew to left, indicating there are outliers that have much high values in the datasets. Compared with the first plot, we can see that the outliers are around 1980.

Consider the variances of the first half and second half of the data:

	Variance of first half	Variance of second half
Mortgage rate	6.96	1.51
Federal Funds rate	10.52	4.22

From the table, we can see that the two rates all have higher variances in the first half and lower variances in the second half, indicating that the time series exhibit heteroscedasticity.

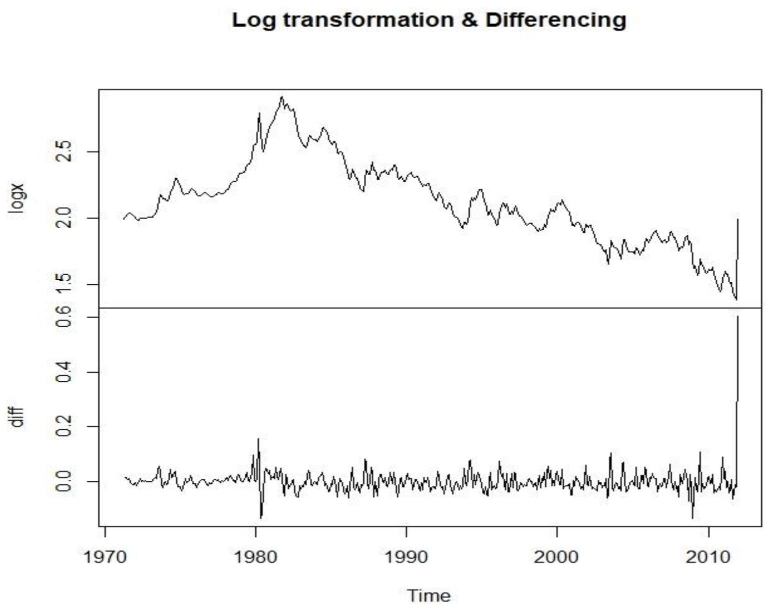
We also consider the autocorrelation of the mortgage rate series:



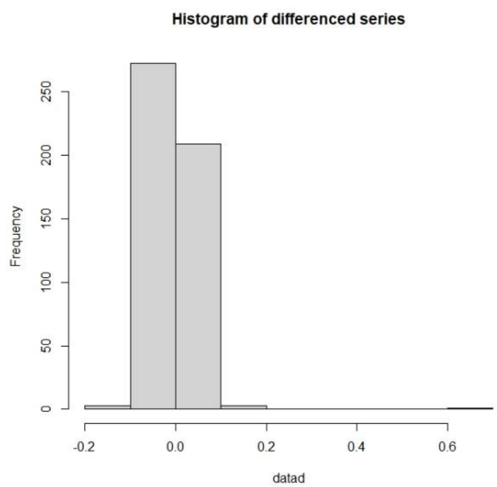
The autocorrelation function (ACF) for a series gives correlations between the series x_t and lagged values of the series. In this plot we see that the ACF tails off slowly, meaning that the correlation between x_t and x_{t-1} is not the same for all t. since we have figured out the mortgage rates have no constant variance, no constant mean, histogram does not follow normal distribution, and the ACF depends on t, we can conclude that the monthly mortgage rate series is not stationary. We need to make proper transformation to make it stationary.

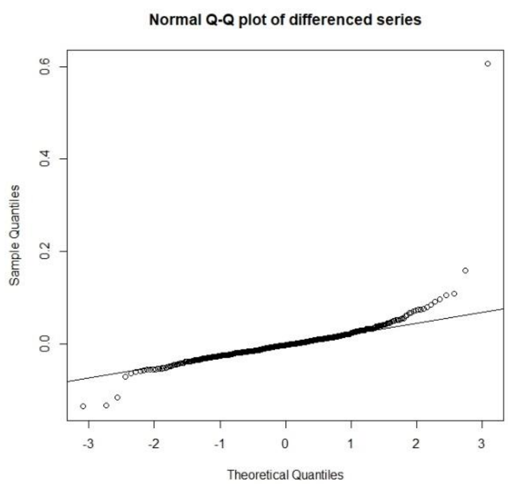
2.2 Methods

We will implement log transformation and differencing to determine the appropriate transformation.



From the plot, we can see that log transformation does not make the series stationary, while differencing seems to make it stationary. We plot the histogram and normal Q-Q plot of the differenced series:



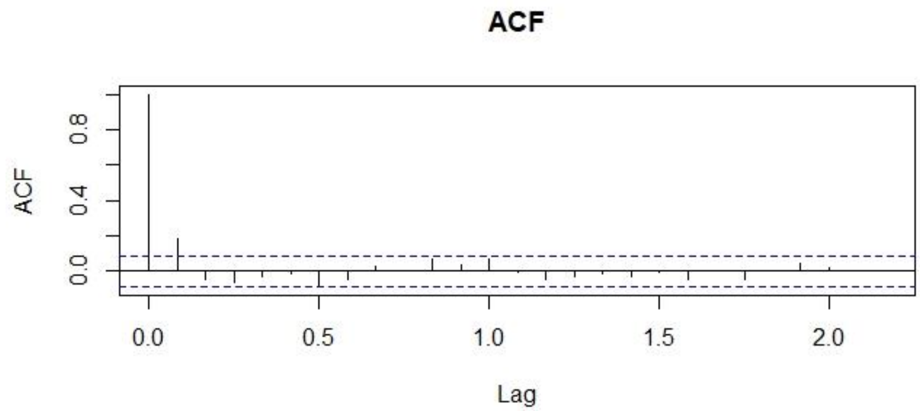


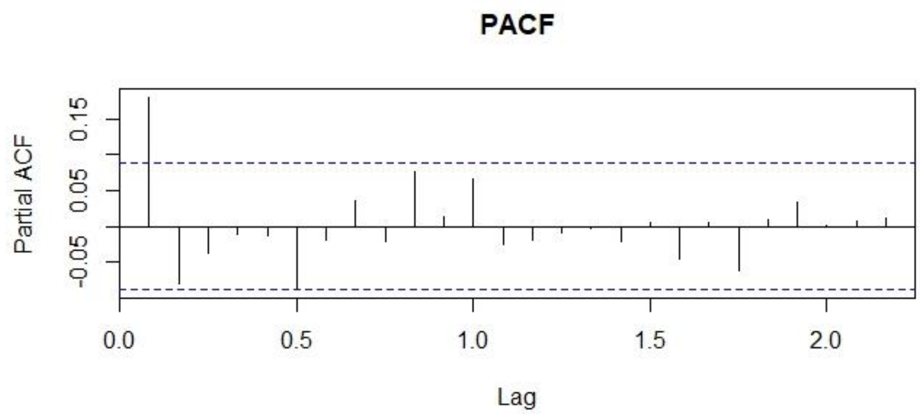
The histogram and the normal Q-Q plot indicate that the transformation improves the normal approximation, thus the differenced log transformation series is stationary.

3 Results

Now we have a stationary series, we want to build an ARIMA model to fit the series. We use the autocorrelation function and the partial autocorrelation function to determine the model.

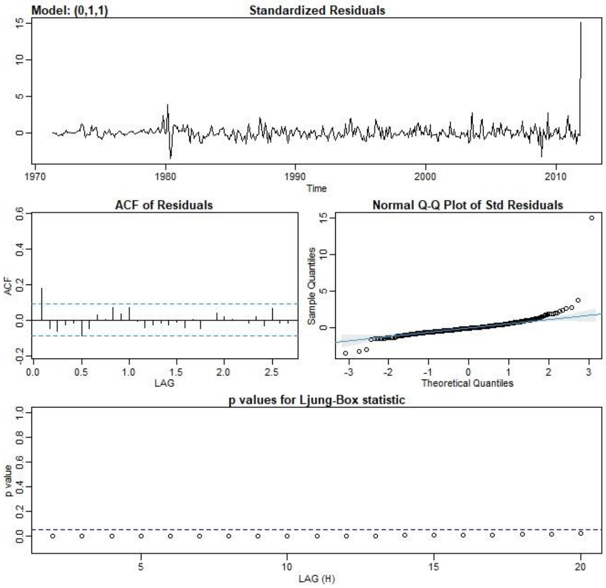
The ACF and PACF is as followed:



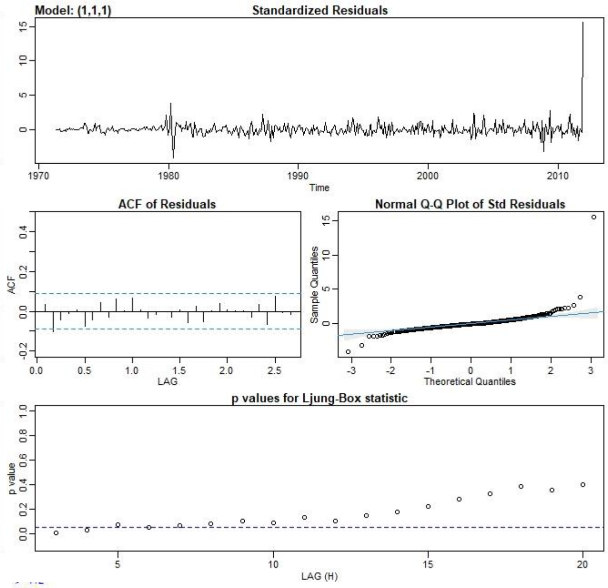


From the plot, we can see that the ACF shows a tendency to cut off after lag 1, but lag 2 is still outside the confidence interval, thus having slightly statistically significantly different from the subsequent autocorrelations. The PACF tails off, showing that the model has infinite AR representation. Therefore, we will check two models, one is ARIMA(1,1,1) and the other is ARIMA(0,1,1).

The result of the two models are:




```
$fit

Call:
stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
    Q), period = S), xreg = constant, transform.pars = trans, fixed = fixed,
    optim.control = list(trace = trc, REPORT = 1, reltol = tol))

Coefficients:
          ma1  constant
      -1.0000         0
s.e.   0.0084         0

sigma^2 estimated as 0.001638:  log likelihood = 867.69,  aic = -1729.38

$degrees_of_freedom
[1] 485

$ttable
         Estimate     SE   t.value p.value
ma1            -1 0.0084 -118.6764   0.000
constant        0 0.0000   -0.1447   0.885

$AIC
[1] -3.551088

$AICc
[1] -3.551037

$BIC
[1] -3.525287
```




```
$fit

Call:
stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
    Q), period = S), xreg = constant, transform.pars = trans, fixed = fixed,
    optim.control = list(trace = trc, REPORT = 1, reltol = tol))

Coefficients:
         ar1      ma1  constant
      0.3384  -1.0000         0
s.e.  0.0606   0.0074         0

sigma^2 estimated as 0.001541:  log likelihood = 883.05,  aic = -1758.09

$degrees_of_freedom
[1] 484

$ttable
         Estimate     SE   t.value p.value
ar1        0.3384 0.0606    5.5835   0.000
ma1       -1.0000 0.0074 -134.9847   0.000
constant   0.0000 0.0000    0.0163   0.987

$AIC
[1] -3.610044

$AICc
[1] -3.609942

$BIC
[1] -3.575644
```

From the standard residual plots, we can see that the residuals of the two models both have constant mean and variances except few outliers. The normal Q-Q plot of the two models are almost the same, and both indicate that the normal distribution provides an adequate fit for the model. The ACF of the residuals of the two models all fall into the 95% confidence interval, indicating that the residuals appear to be random.

The difference of the two models lies in the Ljung-box statistics, AIC and BIC values. The p-values of the Ljung-box statistics for ARIMA(0,1,1) is smaller than the significant level, while the p-value of ARIMA(1,1,1) is higher than the significant level, showing that ARIMA(1,1,1) has more chance to be random and have a better fit. Moreover, comparing the AIC and BIC values, ARIMA(1,1,1) has a smaller AIC and BIC values, indicating that the model has a better fit as well.

Therefore, we choose the ARIMA(1,1,1) model for the mortgage rate series. And the model is:

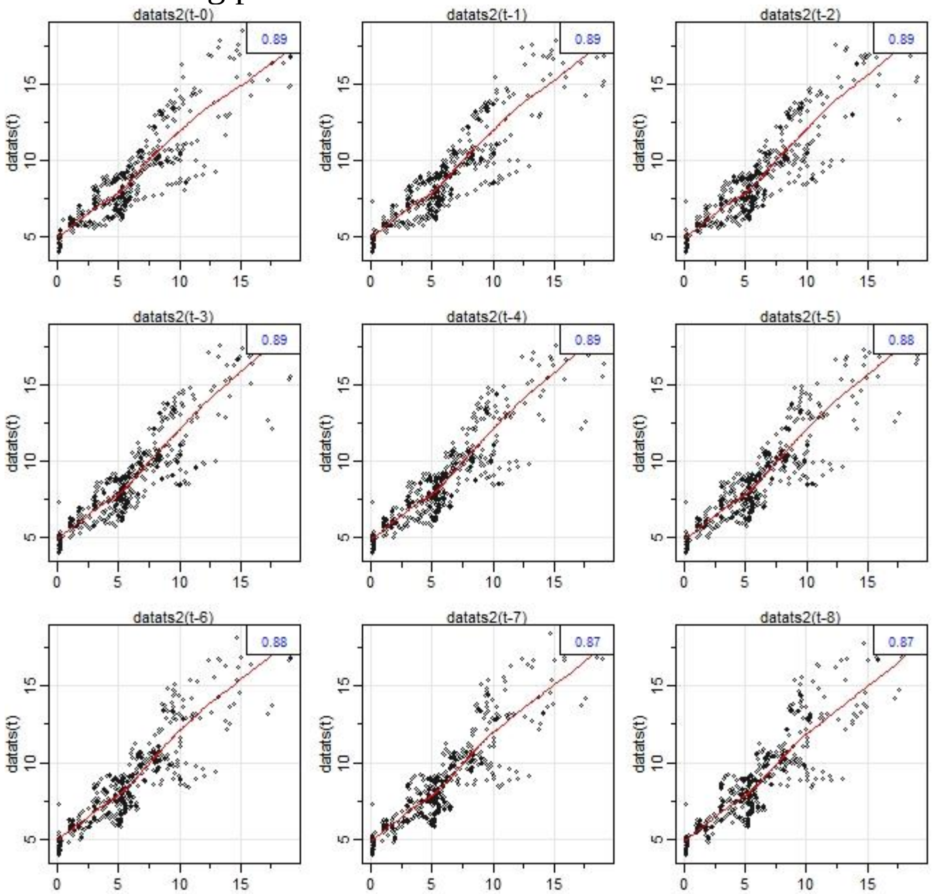
$$X_t = 0.3384X_{t-1} - w_t$$

Now we also want to build a time series model for the mortgage rate using the lag-1 federal funds rate as an explanatory variable. We first assume that the model is:

$$x_t = \beta_0 + \beta_1 F_{t-1} + w_t,$$

where F_{t-1} is the lag-1 federal funds rate and x_t is the mortgage rate.

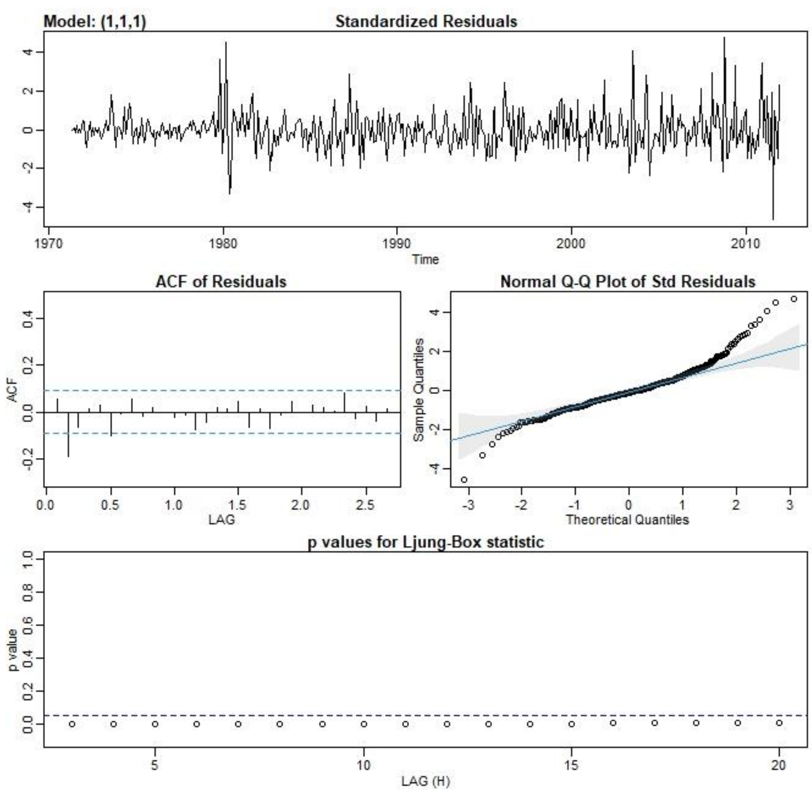
We use the lag plot to check if there exist nonlinear relationships:



The y-axis is the mortgage rate series and the x-axis is the federal funds rate series. From the plot, it can be seen that there is a strong linear relationship within the two variables and no sign of nonlinearity.

Since we have already fit the mortgage rate series with ARIMA(1,1,1), the model that use lag-1 federal funds rate series as explanatory variable is:

```
Call:
stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
    Q), period = S), xreg = xreg, transform.pars = trans, fixed = fixed, optim.control = list(trace = trc,
    REPORT = 1, reltol = tol))

Coefficients:
         ar1      ma1    xreg
      0.2962  -1.0000  0.1403
s.e.  0.0435   0.0071  0.0063

sigma^2 estimated as 0.0007621:  log likelihood = 1054.36,  aic = -2100.73

$degrees_of_freedom
[1] 484

$ttable
     Estimate     SE   t.value p.value
ar1    0.2962 0.0435    6.8093       0
ma1   -1.0000 0.0071 -141.6368       0
xreg   0.1403 0.0063   22.3090       0

$AIC
[1] -4.31361

$AICc
[1] -4.313508

$BIC
[1] -4.279209
```



The residuals are white and follows normal distribution, and the AIC and BIC is small.

Therefore, the model is:

$$x_t = 0.2962x_{t-1} - w_t + 0.1403F_{t-1}$$

4 Appendix

```
setwd("C:/Users/yhwen/Documents/Fall 2020")
data = read.delim('mortgage.txt',sep = " ",header = TRUE)
data1=data[,c('morg')]
data2 = data[,c('ffr')]
datats = ts(data1,start = c(1971,4),end = c(2011,12),frequency = 12)
datats2 = ts(data2,start = c(1971,4),end = c(2011,12),frequency = 12)
tworates = cbind(Morg = datats,
                 Funds = datats2)
plot(tworates,main = 'Mortgage Rate & Federal Funds Rate')

hist(datats,main = 'Histogram of Mortgage rate')
hist(datats2,main = 'Histogram of Federal Funds rate')
datats_p1 = datats[1:(length(datats)/2)]
datats_p2 = datats[(length(datats)/2+1):length(datats)]
var_p1 = var(datats_p1)
var_p2 = var(datats_p2)
datats2_p1 = datats2[1:(length(datats2)/2)]
datats2_p2 = datats2[(length(datats2)/2+1):length(datats2)]
var2_p1 = var(datats2_p1)
var2_p2 = var(datats2_p2)

acf(datats,main = 'ACF of mortgage rate')
morgr = cbind(
              logx = log(datats),
              diff = diff(log(datats)))
plot(morgr,main = 'Log transformation & Differencing')
datad = diff(log(datats))
hist(datad,main = "Histogram of differenced series")
qqnorm(datad, main = 'Normal Q-Q plot of differenced series')
qqline(datad)
acf(datad,main='ACF')
pacf(datad,main = 'PACF')

M1 = astsa::sarima(datad, 0,1,1)
M2 = astsa::sarima(datad,1,1,1)

astsa::lag2.plot(datad2,datad,8)
datad2 = diff(log(datats2))
fish = ts.intersect(datad,ft1 = lag(datad2,-1),dframe = TRUE)
summary(fit1<-lm(datad~ft1,data = fish, na.action=NULL))
ft1 = lag(datad2,-1)
astsa::sarima(datad,1,1,1,xreg=cbind(ft1))
```