Project 2

Riddhi Barbhaiya

11/23/2021

Introduction

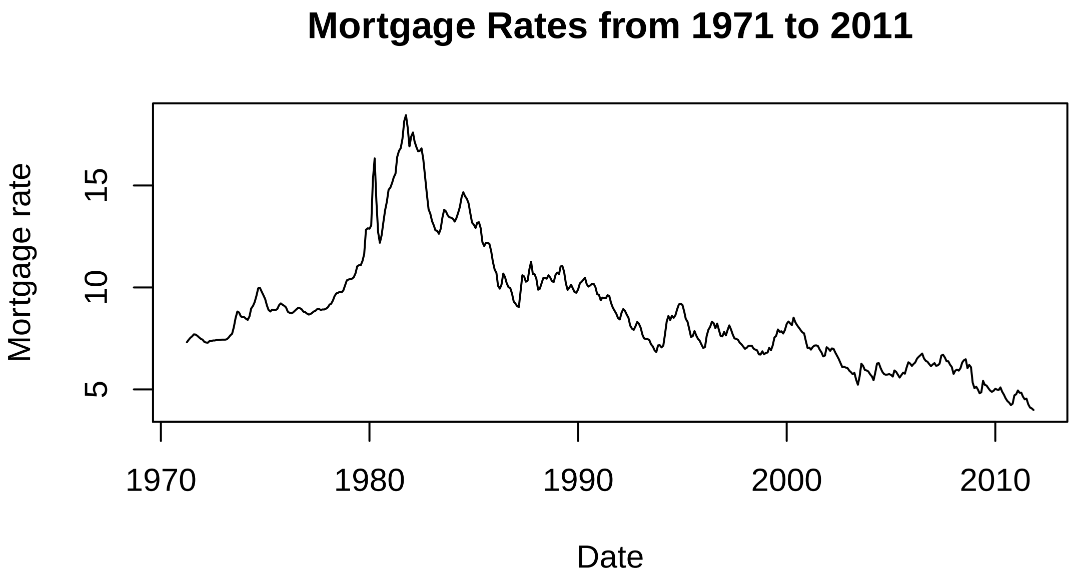
In this project, the monthly mortgage rate from from April 1971 to November 2011 is analyzed and a model is fit to the data. The model may be used to predict future mortgage rates or to explain the relationship of the past rates. First just the monthly mortgage rates are fit to an ARIMA model and then the federeal reserve rate is also considered as an explanatory variable. The report includes the materials and methods section which consists of a descriptions and plots of the data,an analysis of whether the mortgage rates are a stationary series, and transformed rates to satisfy stationarity. Then, using the transformed data, two models are considered, their fit is analyzed, and a model is selected using the Akaike Information Criterion. In the result section, the chosen model is presented and justified. Then, I also present the model which includes the federal reserve rate as an explanatory variable.

Materials and Methods

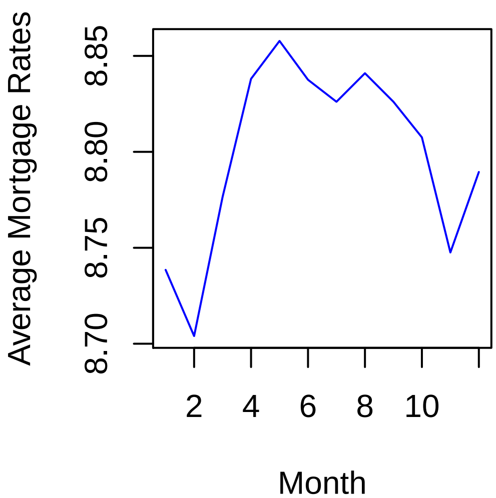
Data

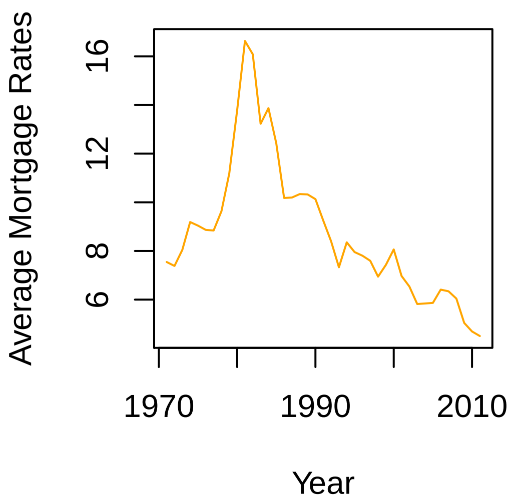
In this project, I consider the US monthly 30-year conventional mortgage rates from April 1971 to November 2011. This data is time series data as per the definition of a time series being a realization of a stochastic process. The mortgage rates are recorded at equally spaced time intervals and the plot below shows how the rates change in time. The data also includes the federal funds rate that is subsequently used as a predictor to fit a model to the mortgage rates.

The following plot shows the mortgage rates over time. The series is not stationary because there are clear trends that such as the rates increasing in the early 1970s and then declining in the late 1970s only to increase more in the 1980s and decline in after the late 1980s.

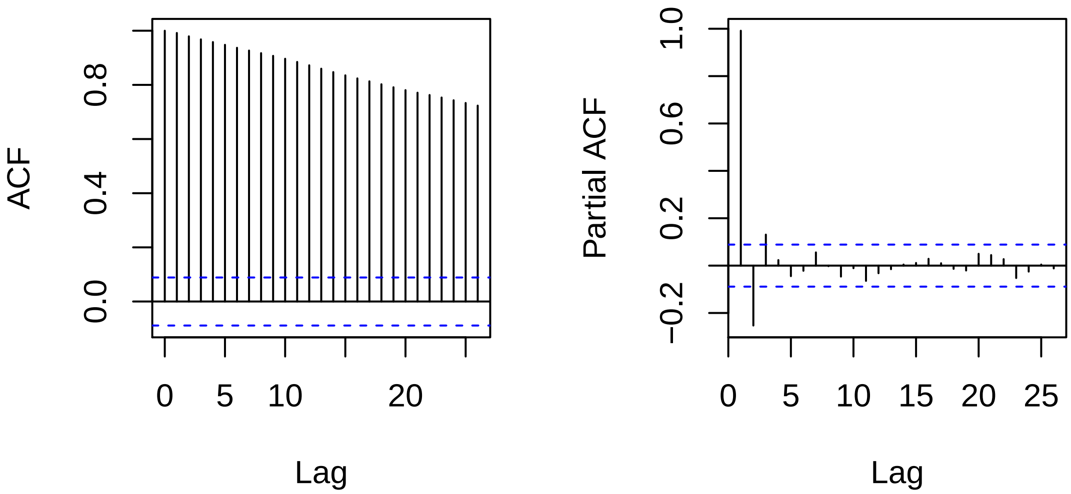


The following plots show the mortgage rates aggregated over month and over the years. These plots show that there isnt a seasonal trend(by month), since the average monthly mortgage rates do not vary systematically. It does make the trend based on year described earlier more appearent.

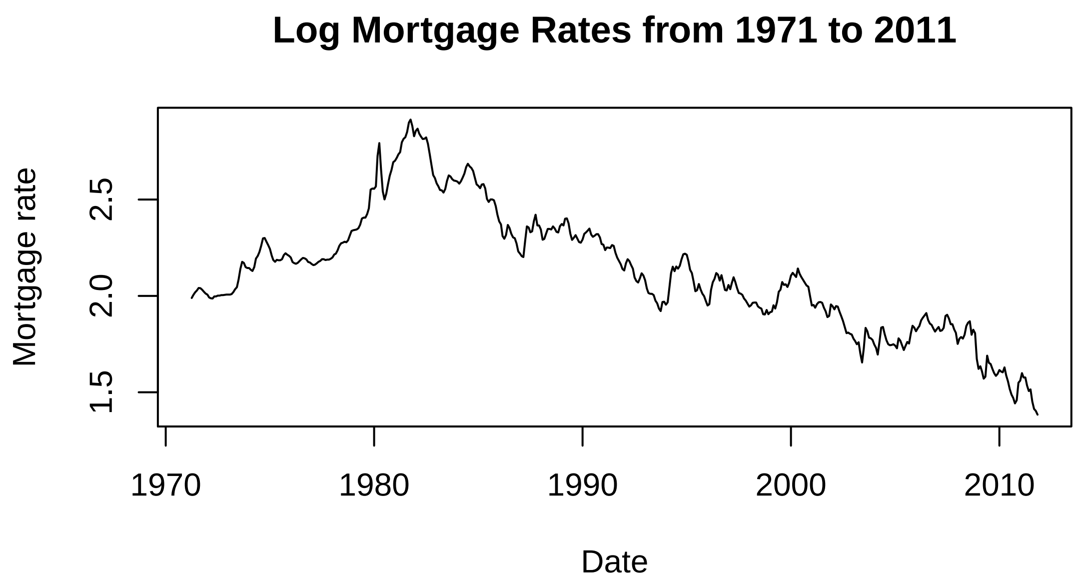




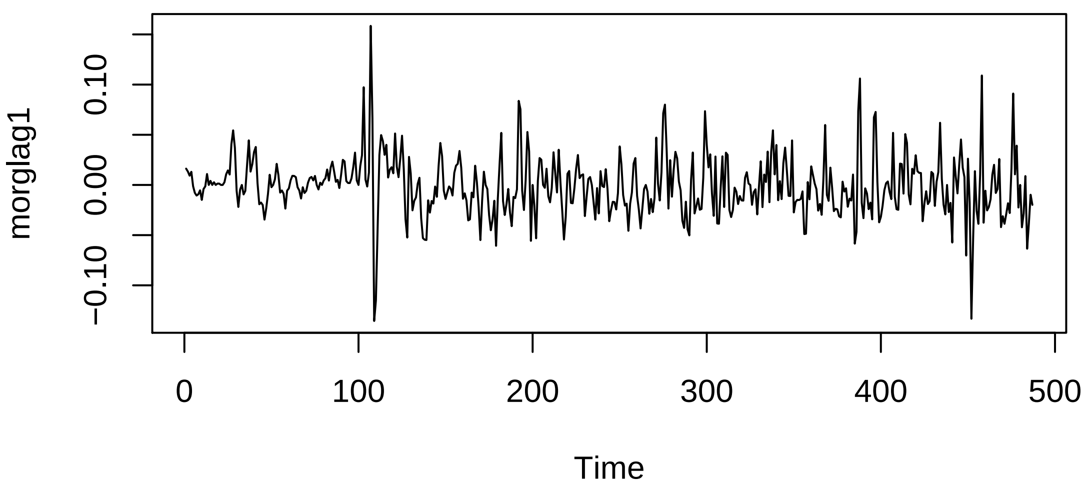
Following I plot the acf and pacf of the series. The acf is decaying very slowly as indicative of a nonstationary series.



As a result, we can say that the timeseries of the mortgage rate is not stationary. To make it stationary, I first apply a log transformation to stabalize the variance.

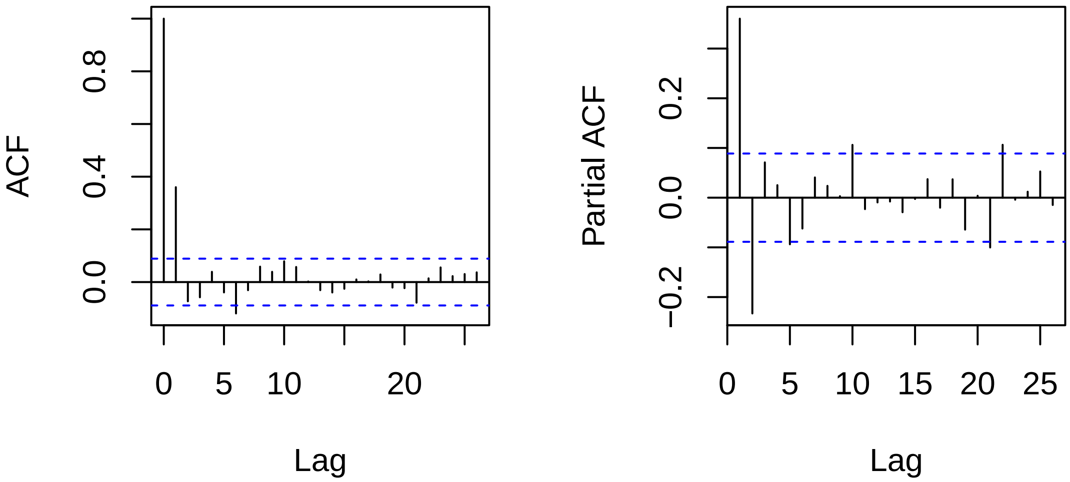


Since this does not eliminate the trend, I also apply lag 1 differencing. While there are still some peaks that do not a resemble a stationary series, the series is close to stationary.



Methods

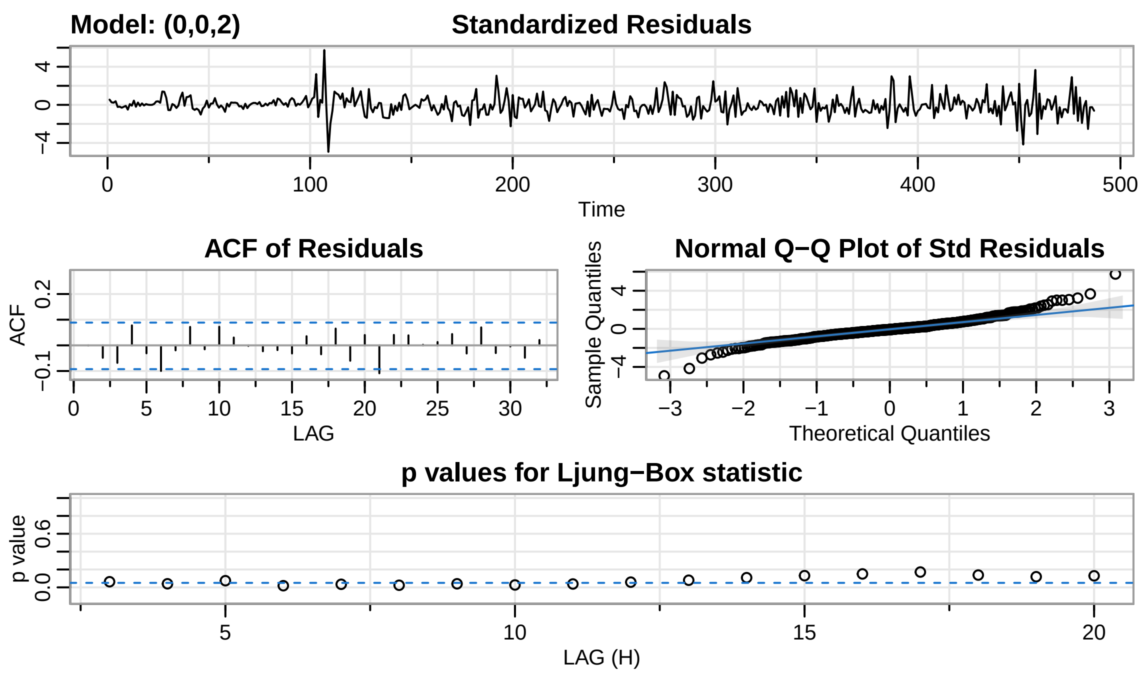
Considering the transformed data above, I plot the acf and pacf to determine possibly apporopriate models for the data.



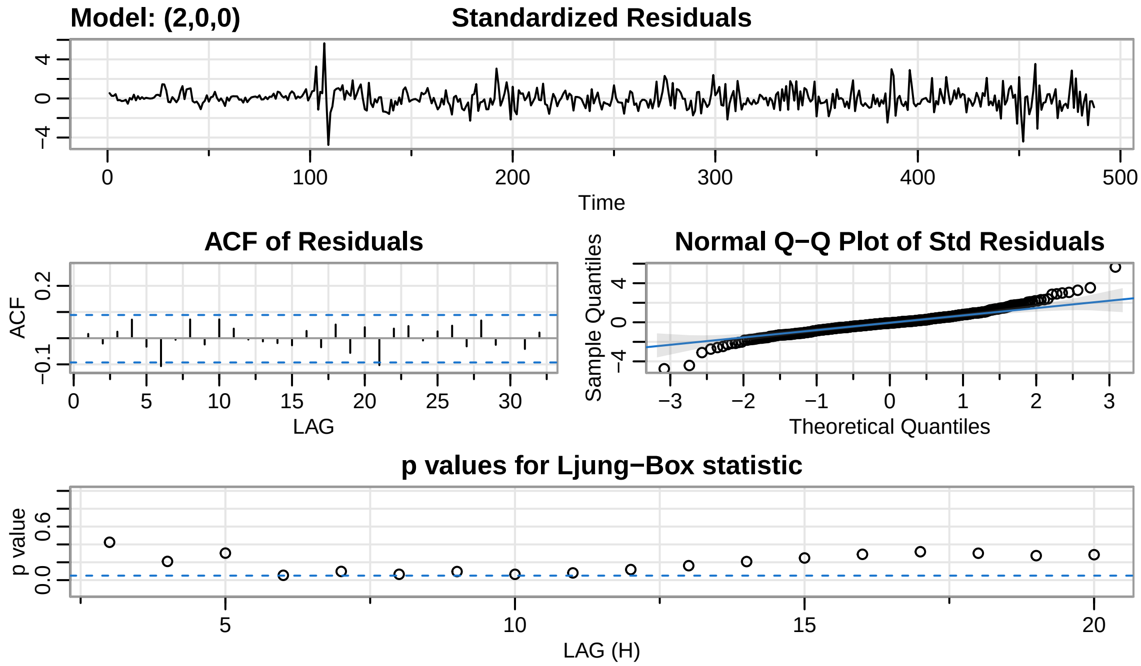
Considering the above plots the following two models are plauible:

- MA(2)- ACF has 2 significant coords and pacf tails off
- AR(2)- Pacf has 2 significant coords and acf tails off

I go on to fit these models using SARIMA. Following is the diagnostic plot for the MA(2) model.



For this model, the standardized residuals resemble white noise(the data varies about 0 with no identifiable trend). There are no signidicant acf values for lag>1, as expected with white noise. The QQ plot indicates that the tails are a bit heavier and that the residuals may not be perfectly normal. Lastly, the p-values for the Ljung-Box statistics are on the border of being significant which indicates that the model may not be a good fit. To compare, I also fit an AR(2) model.



For this model, the diagnostic plots look very similar to the previous model. On thing to not is that the p-values for the Ljung-Box statistics are farther from being significant for more values. Since it is not clear with just the diagnostic plots which model results in the better fit, I compare Akaike information criteria to pick a model.

Table 1: AIC

MA(2)	AR(2)
-4.367909	-4.372796

As seen in Table 1, The AIC is very similar for both models. Therefore, I go with the AR(2) model since it has a better fit as indicated by the p-values for the Ljung-Box statistics. Additionally, in this model both AR coefficients are significant which is not true of the MA(2) model.

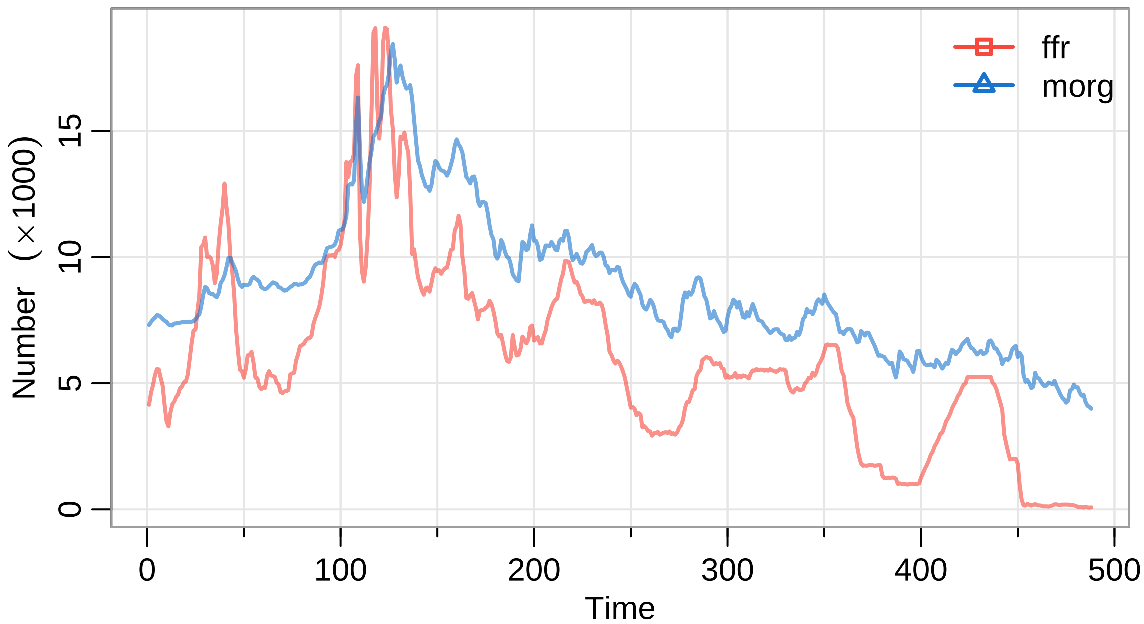
Results

For the reasons mentioned above, I decide to fit a AR(2) model to the lag 1 differenced, log transformed data.Following are the MA coefficients for the model fit using sarima.

Table 2: AR(2) Model Fit

	Estimate	SE	t.value	p.value
ar1	0.4448	0.0441	10.0974	0
ar2	-0.2313	0.0440	-5.2535	0

While this model is satisfactory, we can do better at predicting mortgage rates if we take into account the federal funds rates(ffr). Considering the following plot, we can see that the trend in ffr precedes the mortgage trends. Therefore, I consider the model that uses the lag-1 federal funds rate as an explanatory variable.



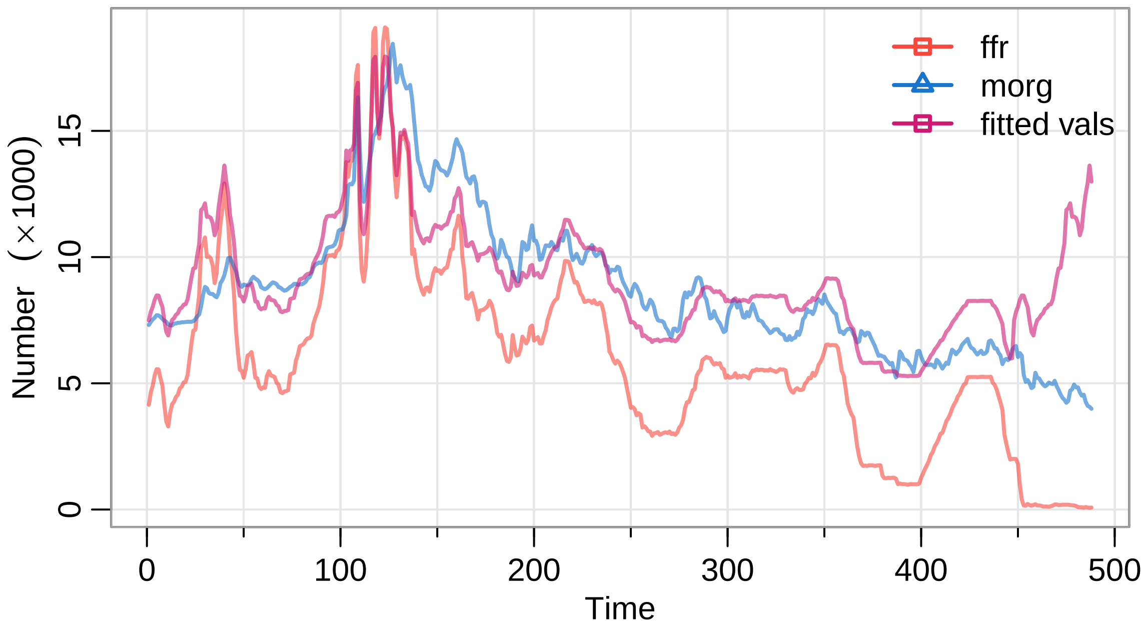
The model considered is as follows:

$$M_t = \beta_0 + \beta_1(F_{t-1}) + x_t$$

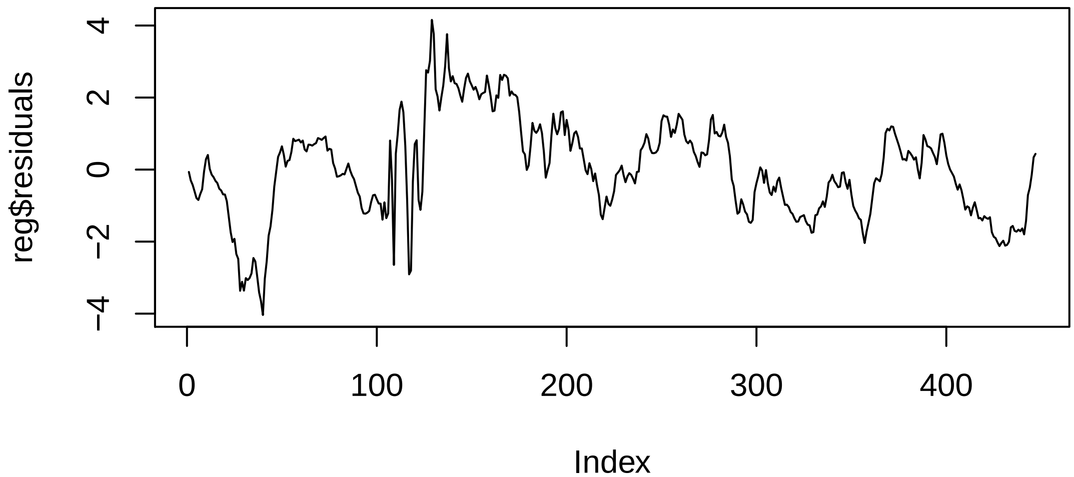
Where M_t is the transformed mortgage rate time series, F_{t-1} is the lag 1 ffr series, and x_t is the autocorrelated error. Following are the fitted estimates. It can be seen that the coefficient for the lag 1 ffr is significant indicating that the ffr is useful in estimating the mortgage rate. Additionaly, we can examine the plot that shows that the fitted values are close to the real mortgage values. Lastly the adjusted r-squared is 0.7616. This indicated that a mojority of the variance in the data is explained by the model.

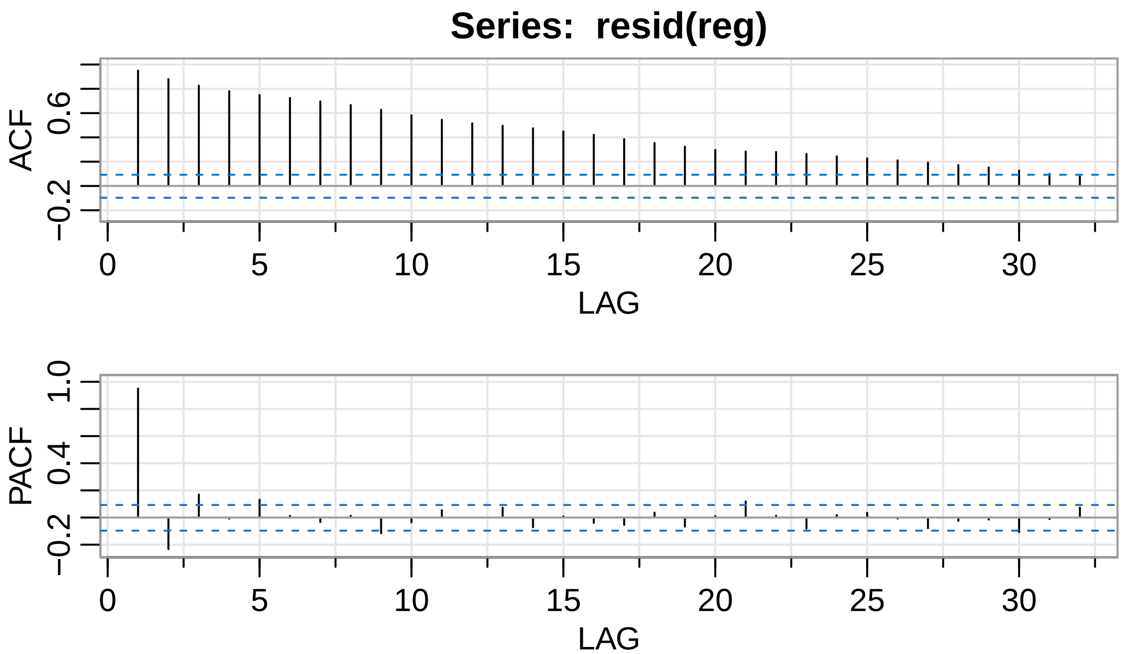
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.593994	0.1366159	33.62709	0
ffrL1	0.699177	0.0185148	37.76321	0

```
## Warning in cbind(data$ffr, data$morg, reg$fitted.values): number of rows of
## result is not a multiple of vector length (arg 3)
```



Lastly, to check the fit of the model, it is necessary to examine the residuals and determine whether they resemble white noise.





These plots show that the residual still has some dependancy(nonzero acf values). Examining just the residual plot, it does not resemble white noise closely. Further modifications must be made to make the residuals look like white noise.

Conclusion

In all, in this project, I examined the monthly mortgage rates from April 1971 to November 2011. To reduce the trends in the data, I transformed and differenced it to make it more stationary. The stationarity of the series could then be used to fit an ARIMA model. I considered two plausible models and picked the better one based on fit and the AIC. Lastly, I regress the lag 1 federal funds rate on monthly mortgage rates. Although this model explains 71% of the variance in mortgage rates, the resulting residual is not white noise alluding to the possibility of further improving the model.

Appendix

```
knitr::opts_chunk$set(echo = TRUE)
knitr::opts_chunk$set(fig.width=6, fig.height=3.5)
library(astsa)
library(kableExtra)
library(dplyr)
#library(zoo)
data = read.delim("/Users/riddhib/Desktop/Fall2021/STA137/Project2/mortgage.txt",
                  header = TRUE, sep = " ")
data$Dates <- as.Date(with(data,paste(data$year,data$month,data$day,sep="-")),"%Y-%m-%d")
df = data.frame(data$Dates, data$morg, data$ffr)
plot(df$data.Dates, df$data.morg, type = 'l', xlab = "Date",
     ylab = "Mortgage rate", main = "Mortgage Rates from 1971 to 2011")

avgmonthly = data %>% group_by(month) %>% summarise(mmg = mean(morg))
avgyearly = data %>% group_by(year) %>% summarise(mmg = mean(morg))
par(mfrow=c(1,2))
```

```
plot(avgmonthly$month, avgmonthly$mmg, type = 'l', col = 'blue', xlab = "Month", ylab = "Average Mortga
plot(avgyearly$year, avgyearly$mmg, type = 'l', col = 'orange', xlab = "Year", ylab = "Average Mortgage
par(mfrow=c(1,2))
acf(df$data.morg, main='')
pacf(df$data.morg, main='')

#par(mfrow=c(1,2))
plot(df$data.Dates, log(df$data.morg), type = 'l', xlab = "Date",
     ylab = "Mortgage rate", main = "Log Mortgage Rates from 1971 to 2011")
#qqnorm(log(df$data.morg), main="", col=4)
 #qqline(log(df$data.morg), col=2, lwd=2)
#par(mfrow=c(1,2))
morglag1 <- diff(log(df$data.morg), 1)
ts.plot(morglag1)
#qqnorm(morglag1, main="", col=4)
 #qqline(morglag1, col=2, lwd=2)

par(mfrow=c(1,2))
acf(morglag1, main='')
pacf(morglag1, main='')
fit1 = sarima(morglag1, p=0, d=0, q=2, no.constant=TRUE)
fit2 = sarima(morglag1, p=2, d=0, q=0, no.constant=TRUE)
a = cbind(fit1$AIC, fit2$AIC)
colnames(a) = c("MA(2)", "AR(2)")
a1 = kable(a, caption = "AIC")
kable_styling(a1, full_width=F, latex_options = "hold_position")
a1 = kable(fit2$ttable, caption = "AR(2) Model Fit")
kable_styling(a1, full_width=F, latex_options = "hold_position")
tsplot(cbind(data$ffr,data$morg), col = astsa.col(c(2,4), .6), lwd=2, type="l", pch=c(0,2),
        spaghetti=TRUE, ylab=expression(Number~~~(""%*% 1000)))
legend("topright", col=c(2,4), lty=1, lwd=2, pch=c(0,2), legend=c("ffr", "morg"), bty="n")
ffrL1 = lag(data$ffr,1)[2:448]
morg = data$morg[2:448]
reg <- lm(morg ~ ffrL1)
a1 = kable(coefficients(summary(reg)))
kable_styling(a1, full_width=F, latex_options = "hold_position")

tsplot(cbind(data$ffr,data$morg, reg$fitted.values),
       col = astsa.col(c(2,4,6), .6),
       lwd=2, type="l", pch=c(0,2), spaghetti=TRUE, ylab=expression(Number~~~(""%*% 1000)))
legend("topright", col=c(2,4,6), lty=1, lwd=2, pch=c(0,2),
       legend=c("ffr", "morg","fitted vals"), bty="n")
plot(reg$residuals, type = 'l')
acf2(resid(reg))
```