STA 137: Project 2

Collin Kennedy

11/25/2020

Introduction

I am presented with a dataset of the US monthly 30-year conventional mortgage rates, as well as the federal funds rate, dating from April 1971 to November 2011. The dataset was obtained from FRED, the Federal Reserve's economic database. The federal funds rate is the interest rate that the Federal Open Market Committee sets a target for, and is the rate at which banks borrow and lend their excess reserves. This is the "interest rate" that is often the focal point of financial news, and for good reason, as it impacts practically every interest rate in the economy. While the Federal Reserve does not explicitly dictate the 30-year mortgage rate (or the fed funds rate for that matter), the Federal Reserve does exercise considerable control over the federal funds rate, which indirectly impacts mortgage rates.

"Mortgage rates" themselves are the rate of interest charged on a mortgage, a debt instrument often used to finance the purchase of homes. Unsurprisingly, this rate is also incredibly important to the public, as mortgage rates often serve as an indicator to potential home buyers whether or not it's a good time to buy a house. This motivates our interest in being able to predict the 30-year mortgage rate based on the mortgage rate of previous time period(s).

More broadly though, my objective is to determine whether or not it is appropriate to model the relationship between mortgage rates, the federal funds rate, and time, and if so, fit the most suitable models to represent the processes. My report therefore consists of the three following sections:

A **Materials and Methods** section to delve into the details and specifics of my model selection, as well as my diagnosis of the statistical relationship between the federal funds rate, mortgage rates, and time. I divide this section in to two subsections:

Model 1: Mortgage Rate; where I develop an ARIMA model for the 30 year mortgage rates, without taking into consideration the federal funds rate, and

Model 2: Mortgage Rates and Federal Funds Rate; where I model (ARIMA) mortgage rates as a function of past observations of the federal funds rate, as well as lags of the mortgage rate itself.

A **Results** section to discuss my results as well as their implications.

And lastly, an **Appendix** section to provide the reader with the R code I used to conduct my analysis and to compile this report.

Materials and Methods

This data represents the changes in the 30 year monthly mortgage rate from April 1971 to November 2011. On a more statistical level, this data is time series data because it is a collection of observations at equally spaced time points.

The dataset consists of 5 variables (columns):

-year
-month
-day

-*morg*: monthly mortgage rate
-*ffr*: federal funds rate

Because this is time series data, I felt it was necessary to do some preliminary data exploration and visualization to gain a sense of the overall relationship between the federal funds rate and mortgage rates with time, respectively.

Histogram: Mortgage Rates

Histogram: Federal Funds rate
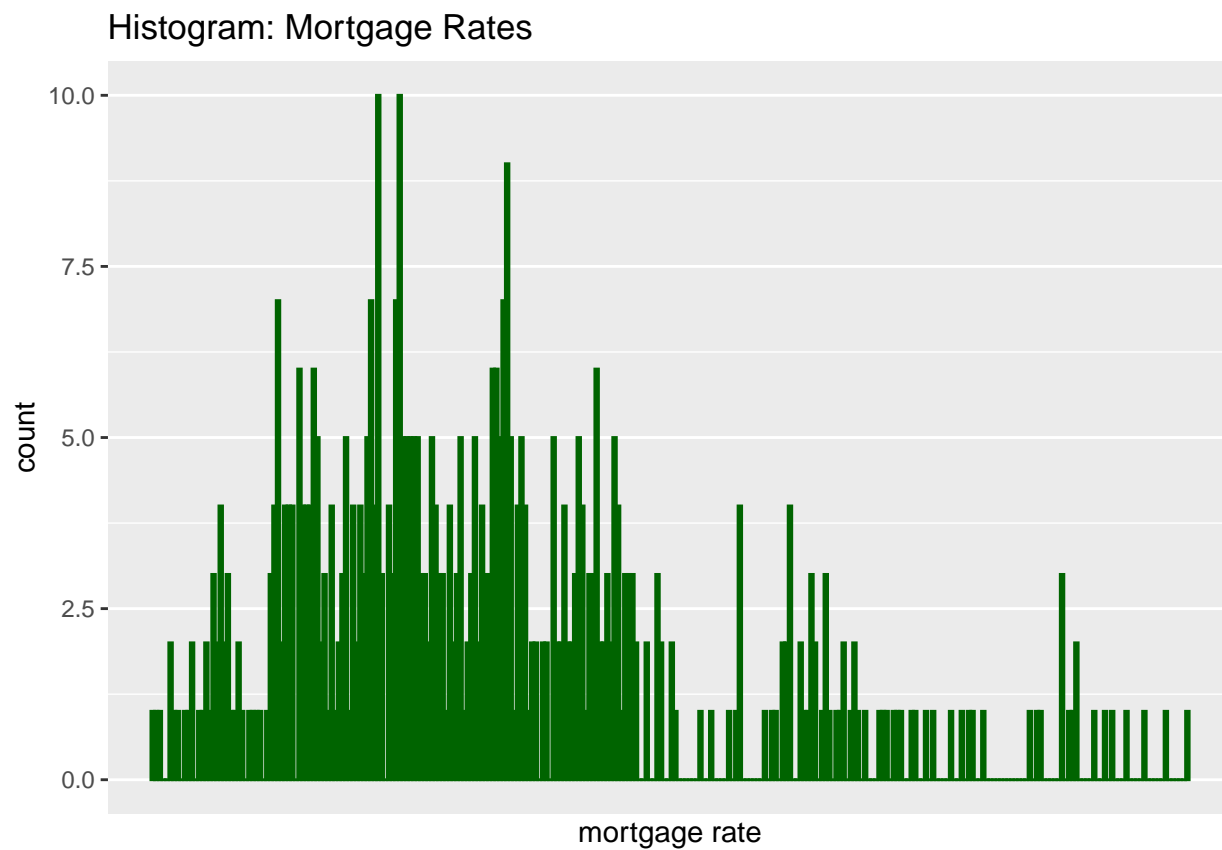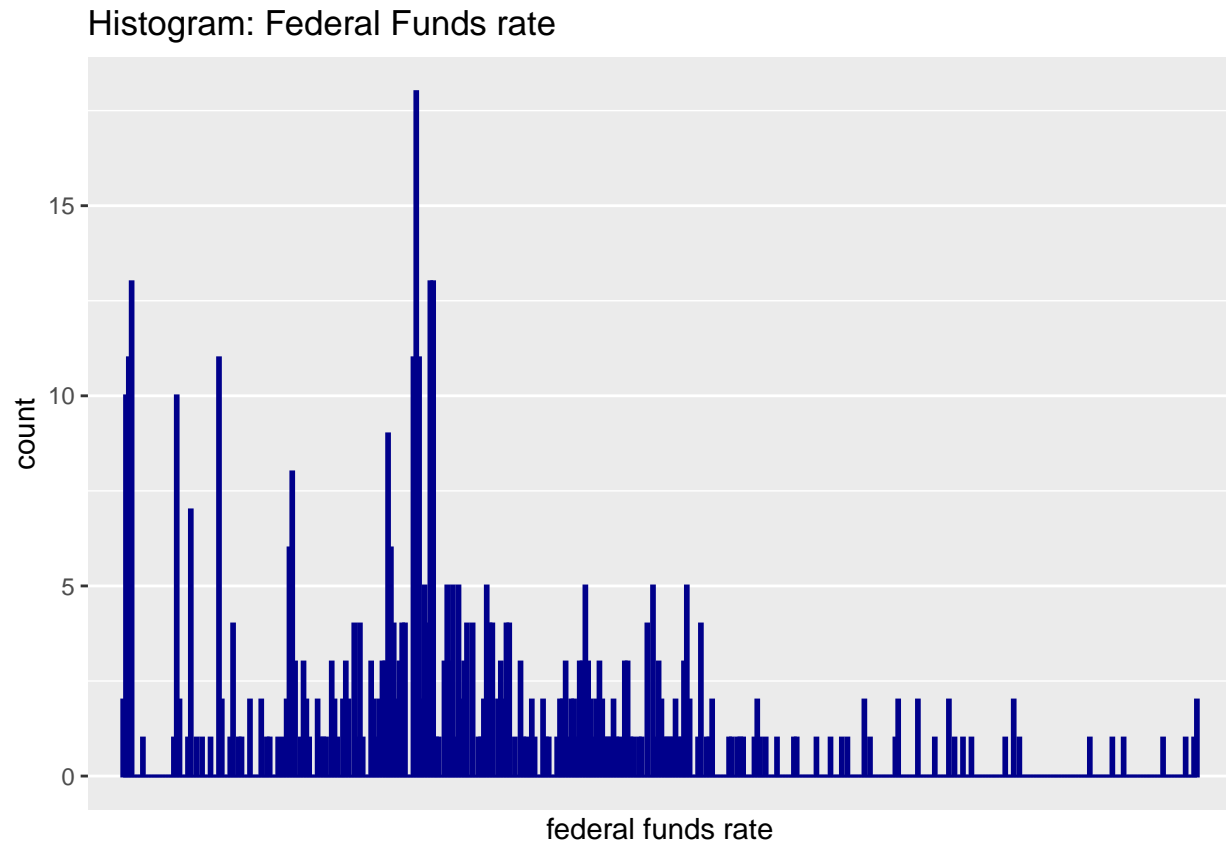
This is notably important for time series data because visualization can help one identify violations of important time series assumptions, such as stationarity.
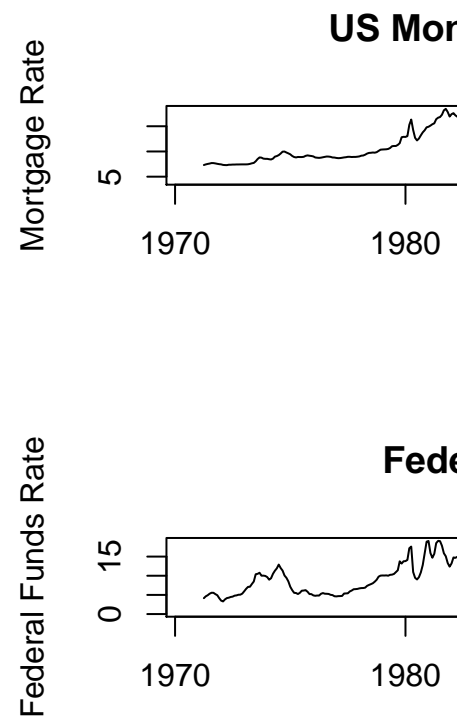
With this in mind, consider the two graphs illustrating the aforementioned relationships:

Given that there is a clear downward trend in the mortgage rate, this is a clear violation of the constant mean assumption for stationarity. As such, I felt it was necessary to perform a transformation to ensure the data was stationary.

I considered a handful of transformations, but ultimately felt that a first-order differencing of the time series
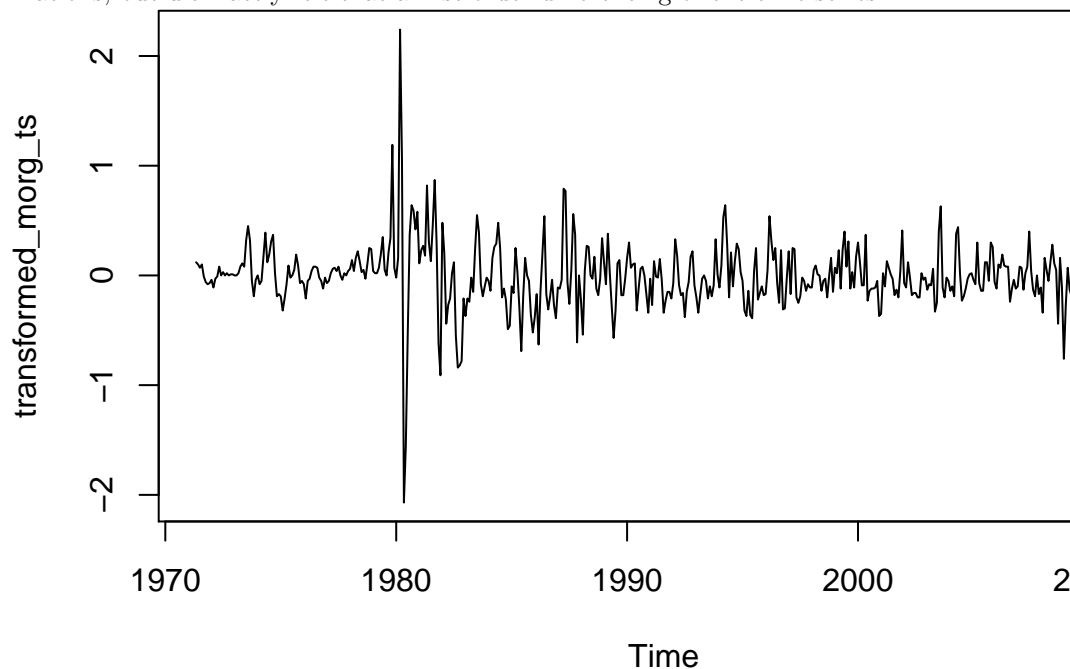

would take care of the time trend:

The differencing does appear to take care of the time trend, and the time series now appears to be stationary. There is clearly an outlier around 1980 (the famous Volcker Recession), but it shouldn't be too problematic going forward.

Next, I consider the autocorrelation and partial autocorrelation of the *transformed* mortgage time series. This will (potentially) allow me to gain a sense of what model(s) are suitable.

```
par(mfrow= c(2,1))
acf(transformed_morg_ts)
pacf(transformed_morg_ts)
```
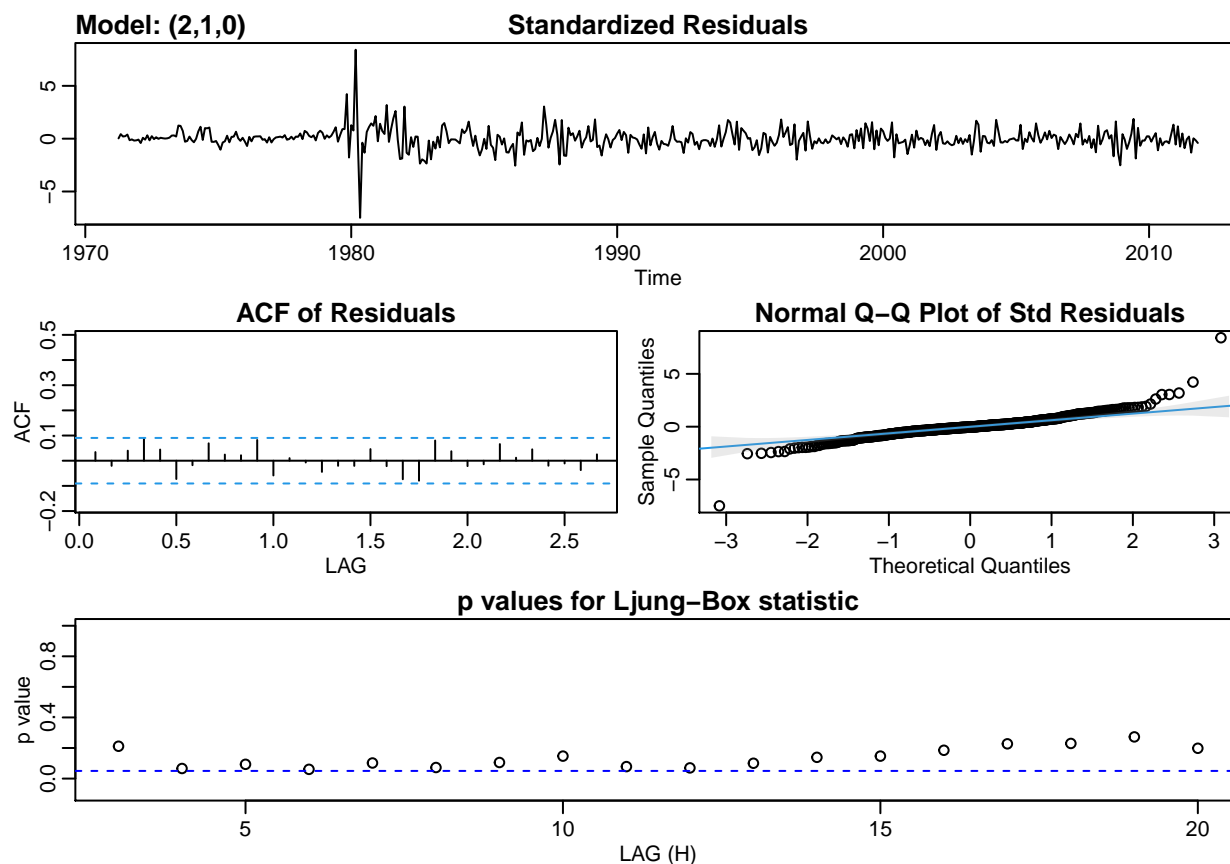
4

Series transformed_morg_ts

Series transformed_morg_ts
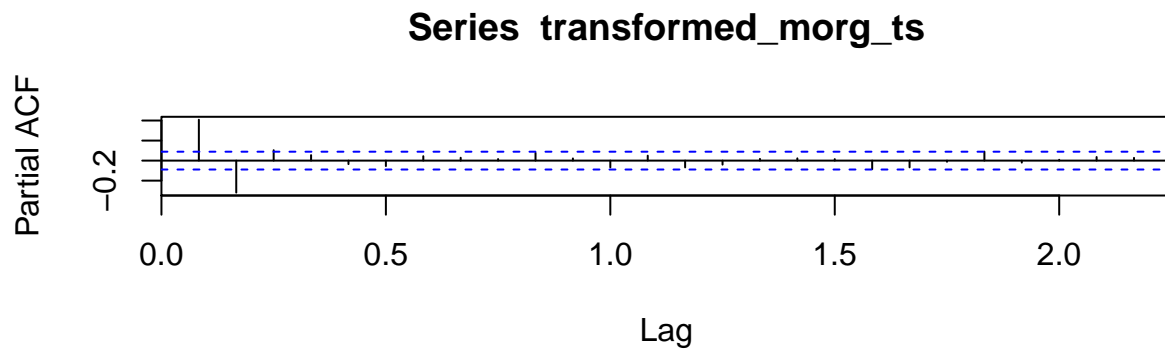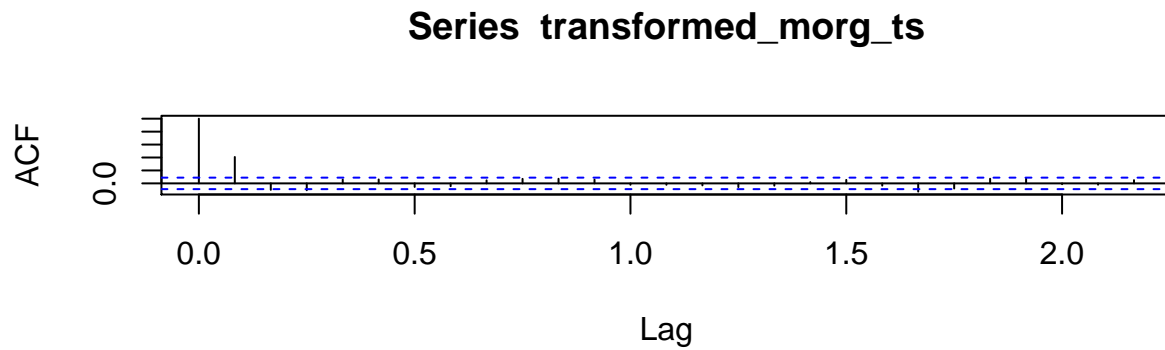
Since both the acf has 1 significant lag and the pacf has 2 significant lags, I could see an Ar(2) or an MA(1) model fitting well.

Model 1: Mortgage Rate

#e In R, I use the sarima() function to construct these two different ARIMA models, and compare them based on the Bayesian Information Criteria. Note that for both models I apply differencing when creating the model, as opposed to differencing the data itself, and then fitting a standard AR(2) or MA(1) to the differenced data. The end result is effectively the same, so it doesn't matter which method I chose. So, I decided to just use *integrated* ARMA (ARIMA) models. First, I consider the ARIMA(2,1,0):

```
sar1 = sarima(mortgage_ts,p = 2,d = 1, q = 0)
```

Model: (2,1,0) — Standardized Residuals / ACF of Residuals / Normal Q–Q Plot of Std Residuals / p values for Ljung–Box statistic
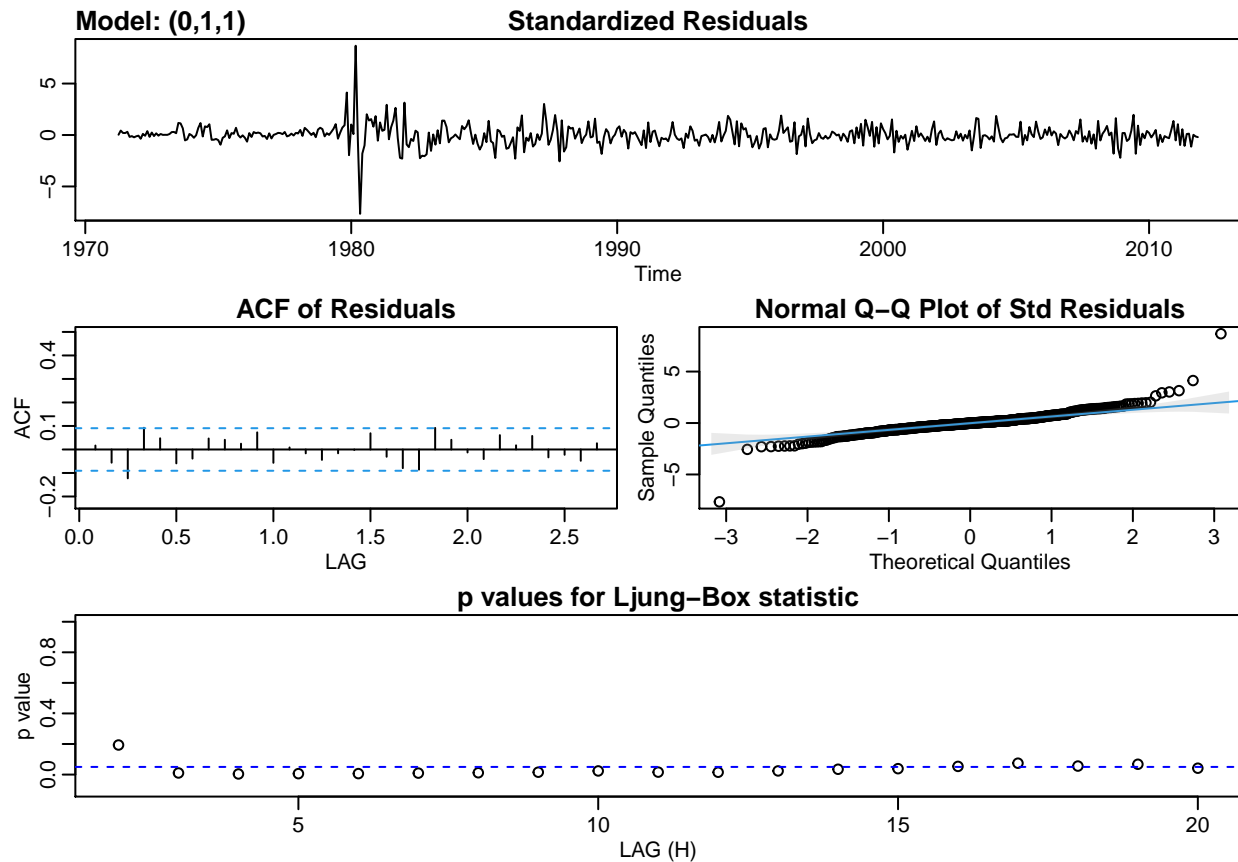
```
sar1$BIC
```

```
## [1] 0.1623507
```

```
sar1$ttable
```

```
##          Estimate      SE t.value p.value
## ar1        0.5378 0.0429 12.5361   0.000
## ar2       -0.3182 0.0429 -7.4251   0.000
## constant  -0.0068 0.0149 -0.4610   0.645
```

Next the *ARIMA(0,1,1)*:

```
sar2 = sarima(mortgage_ts,p=0,d=1,q=1)
```

Model: (0,1,1)   **Standardized Residuals**

ACF of Residuals   **Normal Q–Q Plot of Std Residuals**

p values for Ljung–Box statistic

```
sar2$BIC
```

```
## [1] 0.1638866
```

**Model 2: Mortgage Rates and Federal Funds Rate**

Lastly, because Mortgage rates are known to depend on the Federal Funds rate, I felt it was necessary to construct a model that represents this relationship.

```
laggedFFR = stats::lag(mortgage$ffr)
morgLM = lm(mortgage$morg ~ lag(mortgage$ffr))
summary(morgLM)
```

```
##
## Call:
## lm(formula = mortgage$morg ~ lag(mortgage$ffr))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0241 -0.8486 -0.0180  0.7939  4.1687
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        4.62099    0.11181   41.33   <2e-16 ***
## lag(mortgage$ffr)  0.69606    0.01581   44.02   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 1.301 on 485 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.7998, Adjusted R-squared:  0.7994
## F-statistic:  1938 on 1 and 485 DF,  p-value: < 2.2e-16
```

```
resids = resid(morgLM)
```

Series: resid(morgLM)
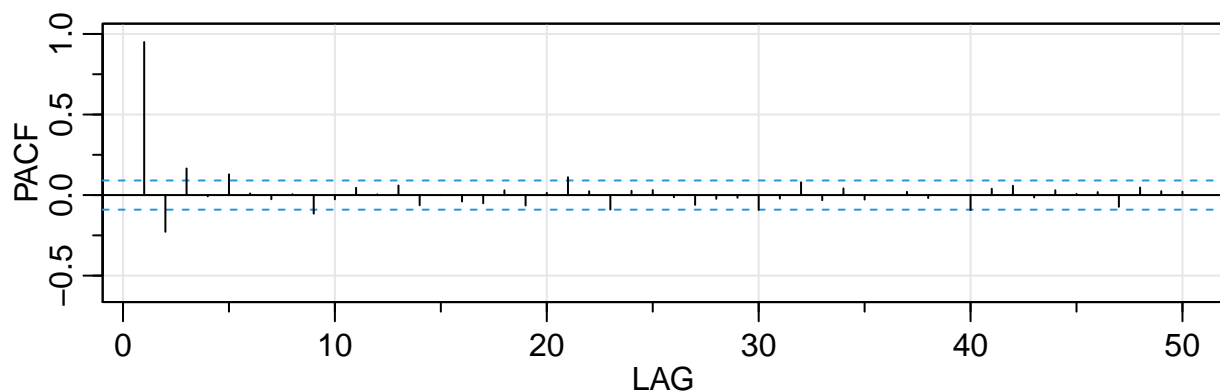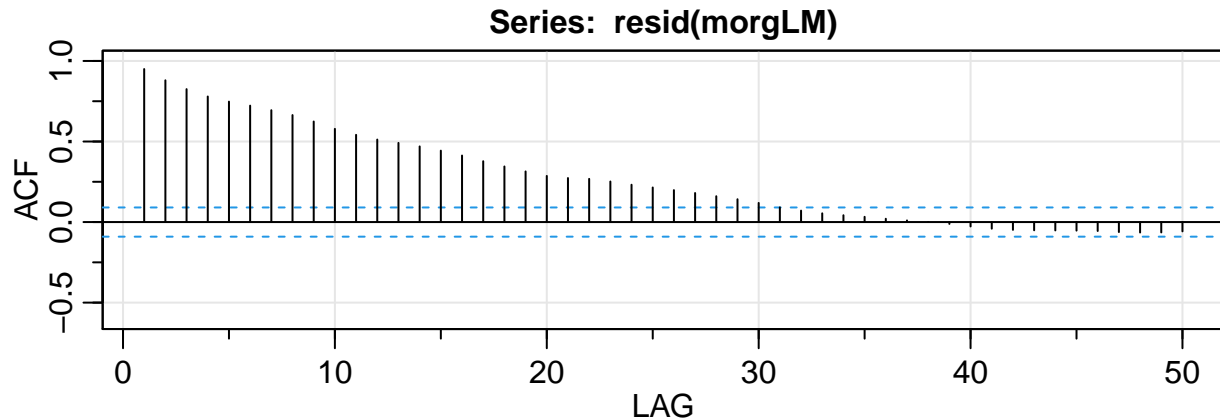


```
##        [,1]   [,2] [,3]   [,4] [,5] [,6]   [,7] [,8]   [,9] [,10] [,11] [,12] [,13]
## ACF   0.95   0.88 0.83   0.78 0.75 0.72   0.69 0.66   0.62  0.58  0.54  0.51  0.49
## PACF  0.95  -0.23 0.17  -0.01 0.13 0.01  -0.03 0.01  -0.12 -0.03  0.05  0.00  0.06
##       [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24] [,25]
## ACF    0.47  0.44  0.41  0.38  0.35  0.31  0.29  0.27  0.27  0.25  0.23  0.21
## PACF  -0.06  0.00 -0.04 -0.05  0.03 -0.06  0.01  0.11  0.02 -0.09  0.03  0.03
##       [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35] [,36] [,37]
## ACF    0.20  0.18  0.16  0.14  0.12  0.09  0.07  0.05  0.04  0.03  0.02  0.01
## PACF  -0.01 -0.06 -0.02 -0.02 -0.09 -0.02  0.08 -0.03  0.04 -0.03  0.00  0.02
##       [,38] [,39] [,40] [,41] [,42] [,43] [,44] [,45] [,46] [,47] [,48] [,49]
## ACF    0.00 -0.01 -0.03 -0.04 -0.05 -0.05 -0.05 -0.05 -0.06 -0.06 -0.06 -0.06
## PACF  -0.02  0.00 -0.09  0.04  0.06 -0.01  0.03  0.01  0.02 -0.07  0.05  0.02
##       [,50]
## ACF   -0.06
## PACF   0.02
```

While the federal funds rate does appear to explain much of the variation in the mortgage rate, it is also evident, based on the ACF plot, that there is significant autocorrelation. I sought to adjust for this serial correlation by modelling the error term as several different ARMA processes, and then selecting the best one again based on BIC:

Fitting the residuals with an AR(1):

```
sarima1 = sarima(mortgage_ts, 1, 0, 0, xreg = cbind(laggedFFR))
```
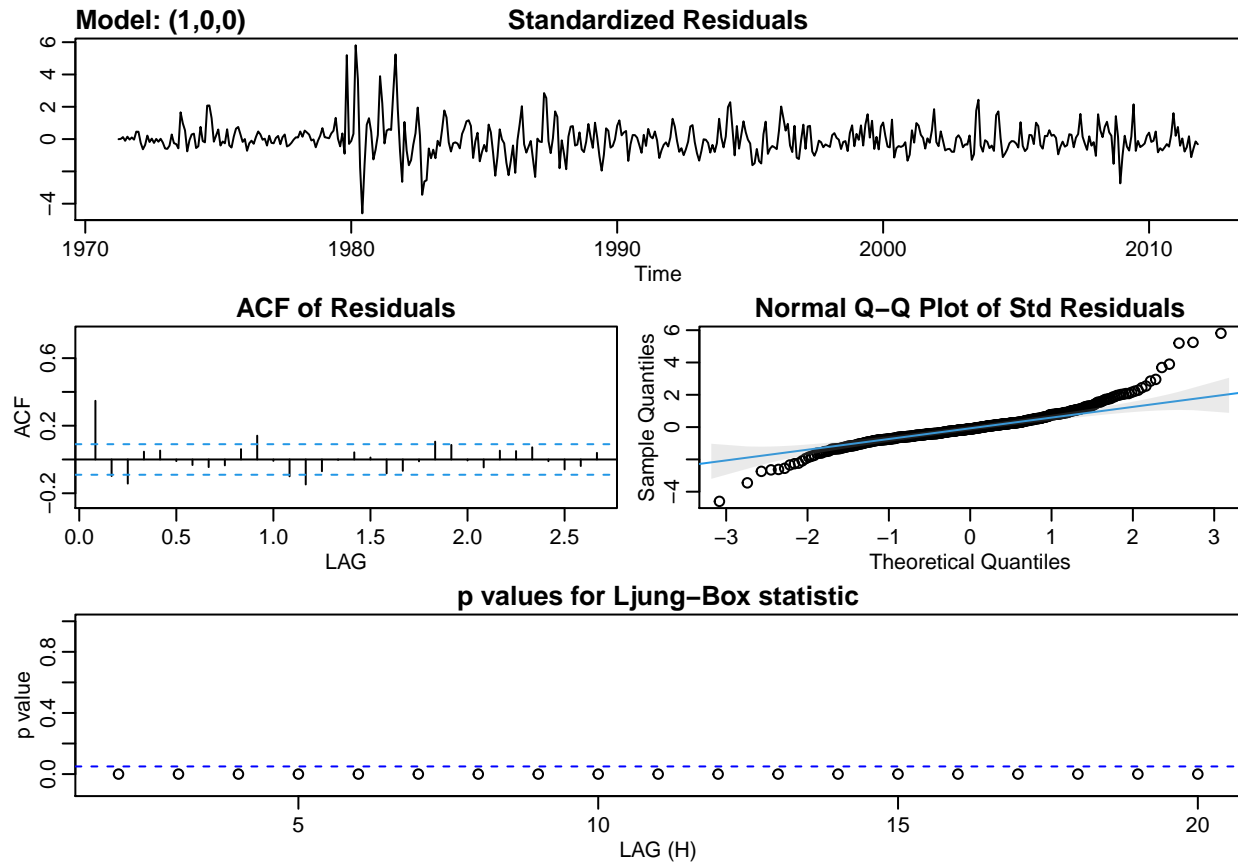

Fitting the residuals with AR(2):

```
sarima2 = sarima(mortgage_ts, 2, 0, 0,xreg = cbind(laggedFFR))
```

```
## Warning in stats::arima(xdata, order = c(p, d, q), seasonal = list(order =
## c(P, : possible convergence problem: optim gave code = 1
```

Model: (2,0,0) — Standardized Residuals / ACF of Residuals / Normal Q–Q Plot of Std Residuals / p values for Ljung–Box statistic
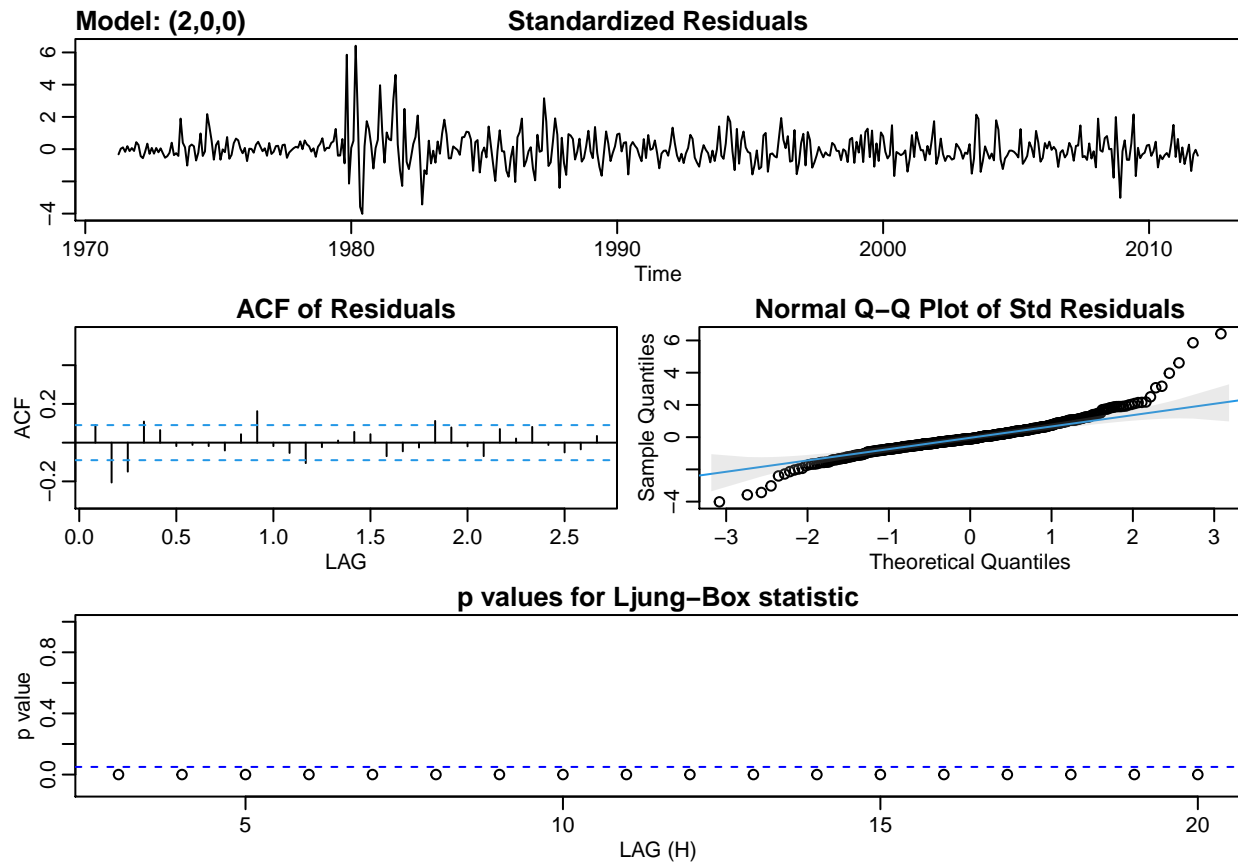
Fitting the residuals with AR(3):

```
sarima3 = sarima(mortgage_ts,3,0,0, xreg = cbind(laggedFFR))
```

Model: (3,0,0)   **Standardized Residuals**

ACF of Residuals   **Normal Q–Q Plot of Std Residuals**

p values for Ljung–Box statistic
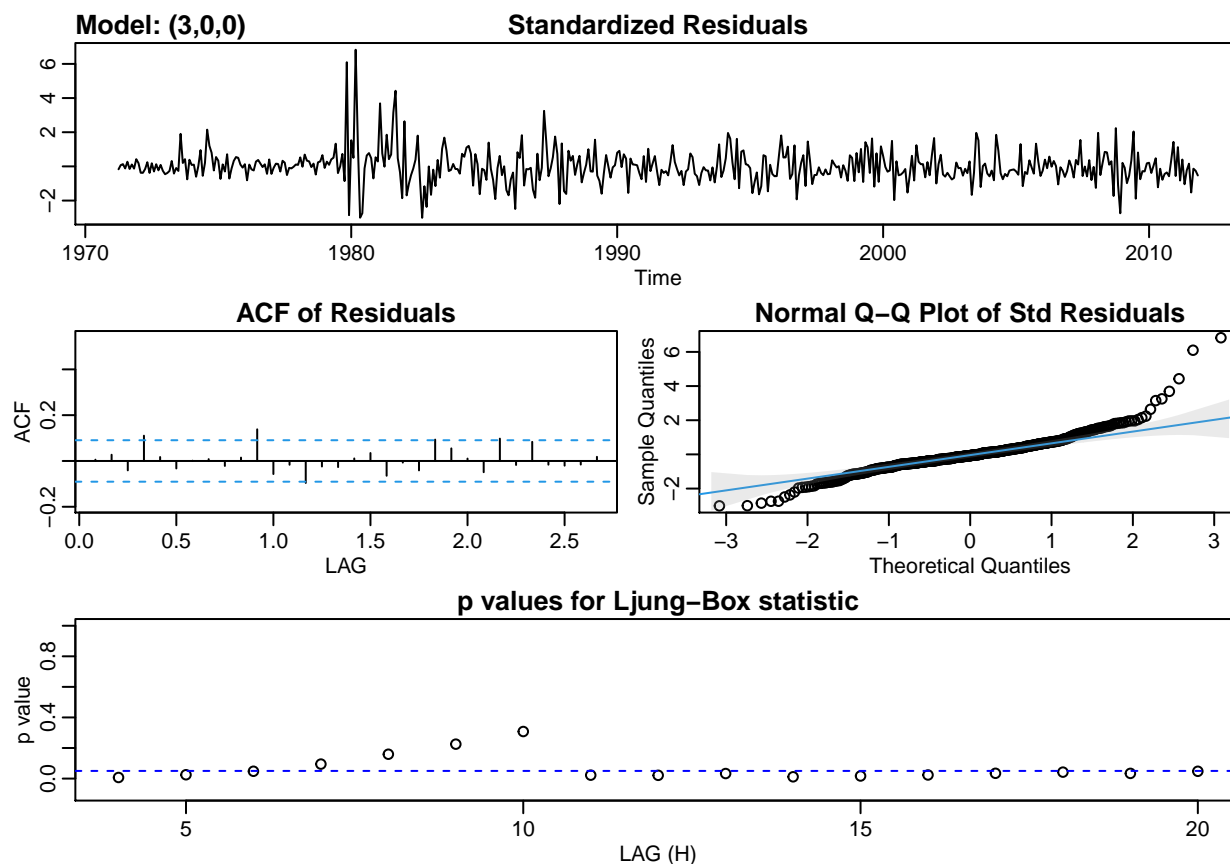
Comparing the fitted coefficients between the 3 possible models:

```
sarima1$ttable
```

```
##             Estimate        SE     t.value p.value
## ar1          1.0000    0.0000 81227.1544  0.0000
## intercept    7.3297  406.7437     0.0180  0.9856
## laggedFFR    0.2445    0.0194    12.5804  0.0000
```

```
sarima2$ttable
```

```
##             Estimate      SE t.value p.value
## ar1           1.3473  0.0426 31.6245       0
## ar2          -0.3564  0.0426 -8.3613       0
## intercept     7.0814  1.0561  6.7050       0
## laggedFFR     0.2268  0.0202 11.2400       0
```

```
sarima3$ttable
```

```
##             Estimate      SE t.value p.value
## ar1           1.4468  0.0448 32.2918       0
## ar2          -0.7117  0.0740 -9.6214       0
## ar3           0.2585  0.0444  5.8234       0
## intercept     6.8517  1.3739  4.9869       0
## laggedFFR     0.2045  0.0202 10.1304       0
```

Now, comparing the Bayesian Information Criteria between the 3 models:

```
sarima1$BIC
```

```
## [1] 0.1711432
```
```
sarima2$BIC
```

```
## [1] 0.05553724
```
```
sarima3$BIC
```

```
## [1] 0.001298539
```

Results

Results: Model 1

Based on the two ARIMA models I fit for the mortgage data, I selected the ARIMA(2,1,0) model:

$$\Delta X_{mortgagerate_t} = 0.5378\Delta X_{mortgagerate_{(t-1)}} - \Delta 0.3182 X_{mortgagerate_{(t-2)}} + w_t$$

First, with the ARIMA(2,1,0) model, the iid white noise assumption appears to remain intact, compared to the ARIMA(0,1,1) model, based on their respective Ljung-Box plots in subsection *Model 1: Mortgage Rates*.

While the difference is marginal, this ARIMA model *does* have a smaller BIC value compared to its ARIMA(0,1,1) counterpart. From this perspective, I figured it was appropriate to choose the ARIMA(2,1,0). BIC gives greater penalty to larger models, and in this case, the ARIMA(2,1,0) model has a smaller BIC while being a *slightly* more complex model.

ARIMA(2,1,0) BIC:

```
## [1] 0.1623507
```

ARIMA(0,1,1) BIC:

```
## [1] 0.1638866
```

In addition to a smaller Bayesian Information Criteria value, AR models are also easier to interpret from an econometric standpoint. In this specific instance, the regression coefficients of the ARIMA(2,1,0) model can be understood as past observations of mortgage rates predicting the current 30-year mortgage rate. This interpretation is more relevant to economists and financiers than say interpreting the coefficients of an MA model(e.g., past error terms/ white noise predicting current mortgage rates).

Results: Model 2

When building the linear model where I regressed the mortgage rate on the *first-order lagged federal funds rate*, there was considerable serial correlation (autocorrelated errors). This serial correlation is cause for concern, as it harms the estimation of regression coefficients. Consider the slow decaying trend in the plot of the autocorrelated errors, as well as the two significant lags in the plot of the PACF:

```
acf2(resid(morgLM),ylim=c(-.6,1),50)
```
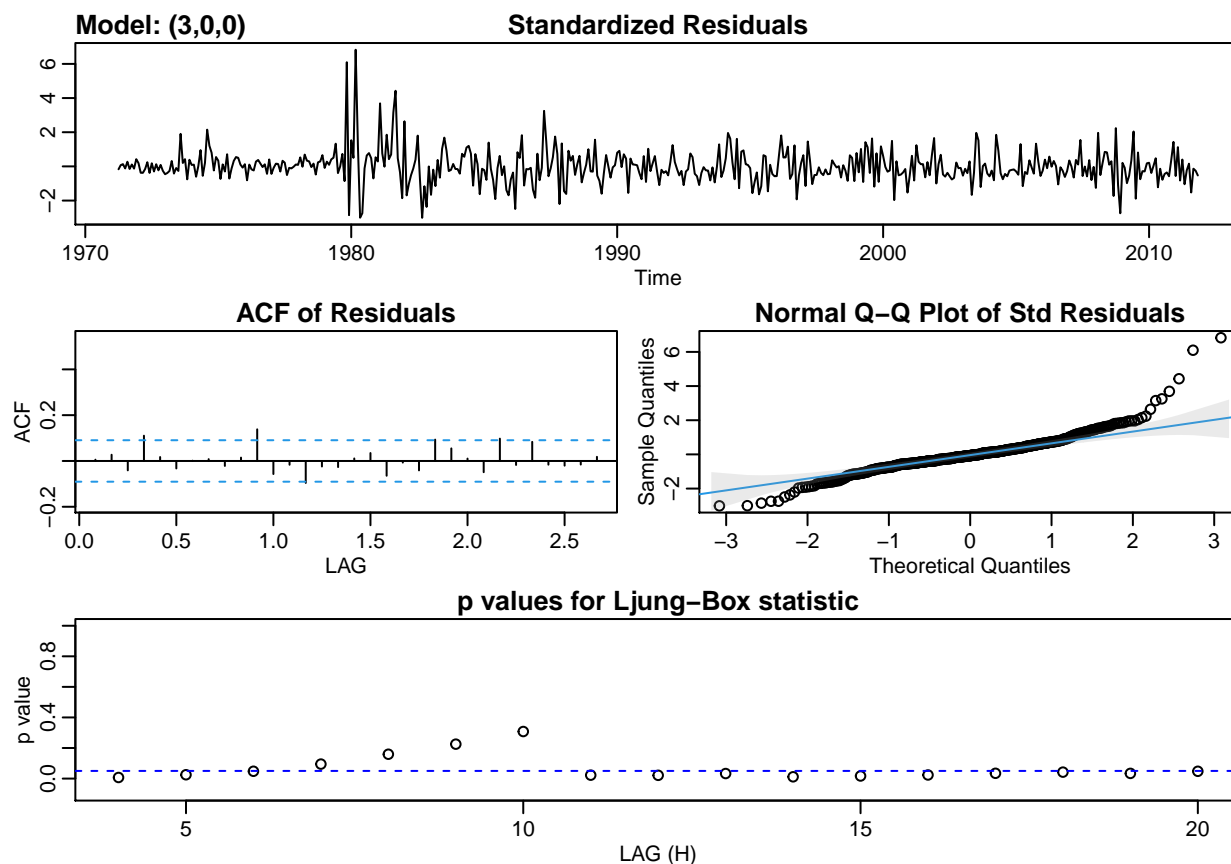
Series: resid(morgLM)
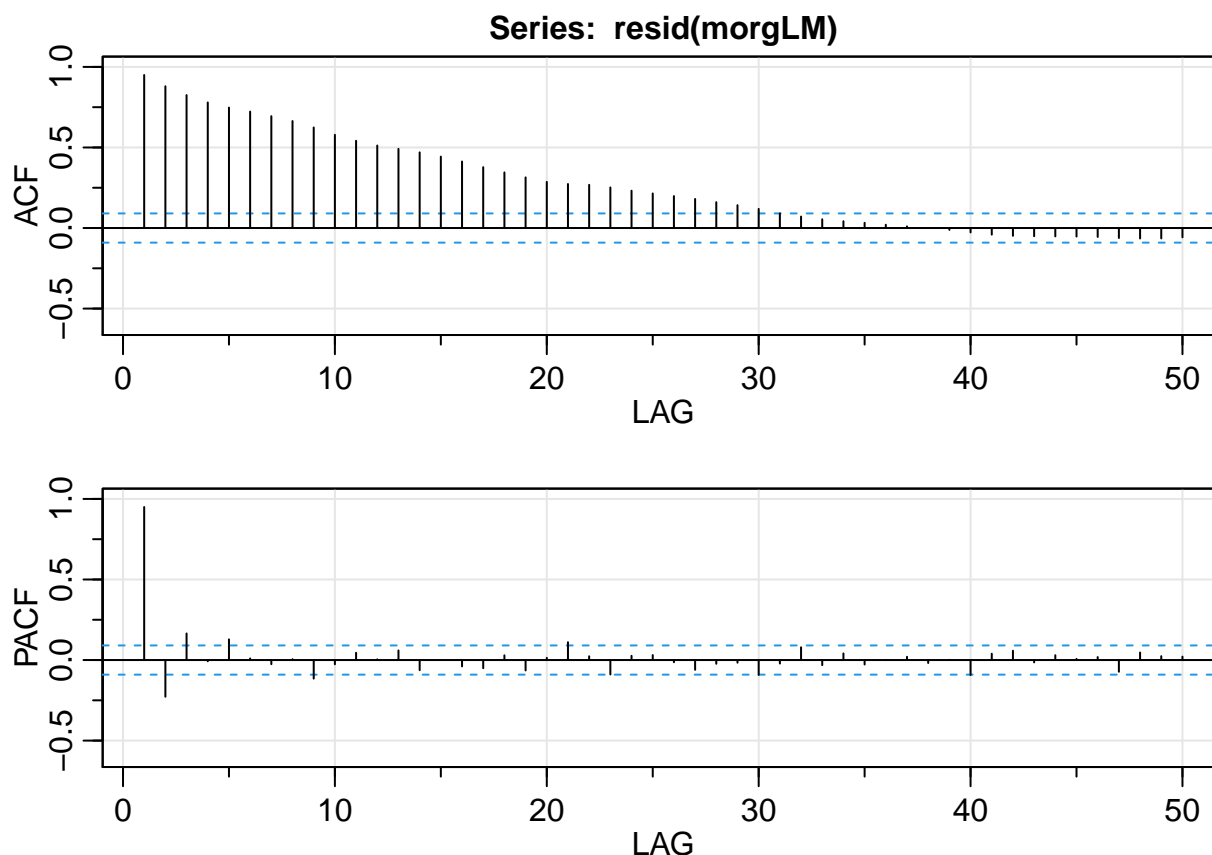
Thankfully, there are methods to account for this serial correlation. The two significant lags indicated to me that an AR(3) process would likely be a suitable model for the serial correlation. Using the sarima() function, I was able to make an adjusted regression, by designating the order of the model for the autocorrelated error, and including my lagged federal funds rate variable as an external regressor:

```
sarima3$fit
```

```
##
## 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 =
##     REPORT = 1, reltol = tol))
##
## Coefficients:
##          ar1      ar2     ar3  intercept  laggedFFR
##       1.4468  -0.7117  0.2585     6.8517     0.2045
## s.e.  0.0448   0.0740  0.0444     1.3739     0.0202
##
## sigma^2 estimated as 0.05379:  log likelihood = 18.25,  aic = -24.51
```

This final ARIMA model allows us to model the relationship between current 30-year mortgage rates, and past mortgage rates and the federal funds rate. Notice how by adjusting for the serial correlation, the errors no longer appear to be autocorrelated. Consider the "ACF of Residuals" plot below:

```
sarima2 = sarima(mortgage_ts, 3, 0, 0,xreg = cbind(laggedFFR))
```

Model: (3,0,0) — **Standardized Residuals**

ACF of Residuals

Normal Q–Q Plot of Std Residuals

p values for Ljung–Box statistic

Notice the slow decay in the ACF of Residuals plot is gone, and that the error at the vast majority of lags are statistically insignificant. This model also had the smallest BIC value compared to the other 2 models I tried in subsection **Model 2: Mortgage Rates and Federal Funds Rate** of my **Materials and Methods Section** (denoted as "AR3_BIC" in the table below):

```
##        AR1_BIC      AR2_BIC      AR3_BIC
## BIC 0.1711432 0.001298539 0.001298539
```

Therefore, my final fitted ARIMA model:

$$\hat{X}_{mortgagerate_t} = 6.8517 + 0.2045 X_{ffr,(t-1)} + 1.4468 X_{mortgagerate,(t-1)} - .7117 X_{mortgagerate,(t-2)} + .2585 X_{mortgagerate,(t-3)}$$

This ARIMA model will allow us to make predictions about the current 30 year mortgage rate based on past (mortgage) rates, as well as the federal funds rate of the previous time period.

Appendix

```r
knitr::opts_chunk$set(echo = TRUE)
library(ggplot2)
library(tidyverse)
library(astsa)
library(lubridate)
mortgage = read.table("/Users/collinkennedy/Downloads/Downloads/mortgage.txt", header = TRUE, sep=" ")

mortgage_ts <- ts(mortgage$morg, start=c(1971, 4), end=c(2011, 11), frequency=12)
plot(mortgage_ts)
```

14

```r
ffr_ts = ts(mortgage$ffr, start=c(1971, 4), end=c(2011, 11), frequency=12)

morg_hist = ggplot(mortgage, aes(x = morg)) +
  geom_histogram(binwidth = 0.05, fill = "darkgreen", color = "darkgreen")+
  scale_x_discrete("mortgage rate")+
  ggtitle("Histogram: Mortgage Rates") #fix x axis

ffr_hist = ggplot(mortgage, aes(x = ffr)) +
  geom_histogram(binwidth = 0.05, fill = "darkgreen", color = "darkblue")+
  scale_x_discrete("federal funds rate")+
  ggtitle("Histogram: Federal Funds rate") #fix x axis

#library("gridExtra")
#grid.arrange(morg_hist,ffr_hist)

par(mfrow = c(2,1))
morg_hist
ffr_hist
par(mfrow= c(2,1))
morgTSPlot = plot.ts(mortgage_ts, main = " US Monthly Mortgage Rate: 1971 -2011",ylab= "Mortgage Rate")
ffrTSPlot = plot.ts(ffr_ts, ylab = "Federal Funds Rate", main = "Federal Funds Rate: 1971 - 2011")


transformed_morg_ts = diff(mortgage_ts,1)
plot(transformed_morg_ts)
par(mfrow= c(2,1))
acf(transformed_morg_ts)
pacf(transformed_morg_ts)
sar1 = sarima(mortgage_ts,p = 2,d = 1, q = 0)
sar1$BIC
sar1$ttable
sar2 = sarima(mortgage_ts,p=0,d=1,q=1)
sar2$BIC


laggedFFR = stats::lag(mortgage$ffr)
morgLM = lm(mortgage$morg ~ lag(mortgage$ffr))
summary(morgLM)
resids = resid(morgLM)


#check the acf of the residuals
acf2(resid(morgLM),ylim=c(-.6,1),50)

sarima1 = sarima(mortgage_ts, 1, 0, 0, xreg = cbind(laggedFFR))
sarima2 = sarima(mortgage_ts, 2, 0, 0,xreg = cbind(laggedFFR))
sarima3 = sarima(mortgage_ts,3,0,0, xreg = cbind(laggedFFR))

sarima1$ttable

sarima2$ttable

sarima3$ttable
```

```
sarima1$BIC

sarima2$BIC

sarima3$BIC


sar1$BIC
sar2$BIC
acf2(resid(morgLM),ylim=c(-.6,1),50)
sarima3$fit


sarima2 = sarima(mortgage_ts, 3, 0, 0,xreg = cbind(laggedFFR))

outputDF = data.frame(AR1_BIC = sarima1$BIC, AR2_BIC = sarima2$BIC, AR3_BIC = sarima3$BIC)
row.names(outputDF) = c("BIC")
outputDF
```