30-year Fixed Rate Mortgage: Average Rate

•••

MA 4780

Ian Boulis and Ella Faulk

04/20/2022

30-Year Fixed Rate Mortgage Background

- Mortgage to be paid off in 30 years where the Interest Rate remains the same over the 30 years

Parts:
- Principal: the original amount borrowed
- Interest: what is paid on top of the principal

Depends on each person:
- Credit Score
- Home price and down payment
- Location
- Loan type (i.e. VA loans)

https://www.rocketmortgage.com/learn/30-year-fixed-mortgage-rate

Why is this data important?

Money!

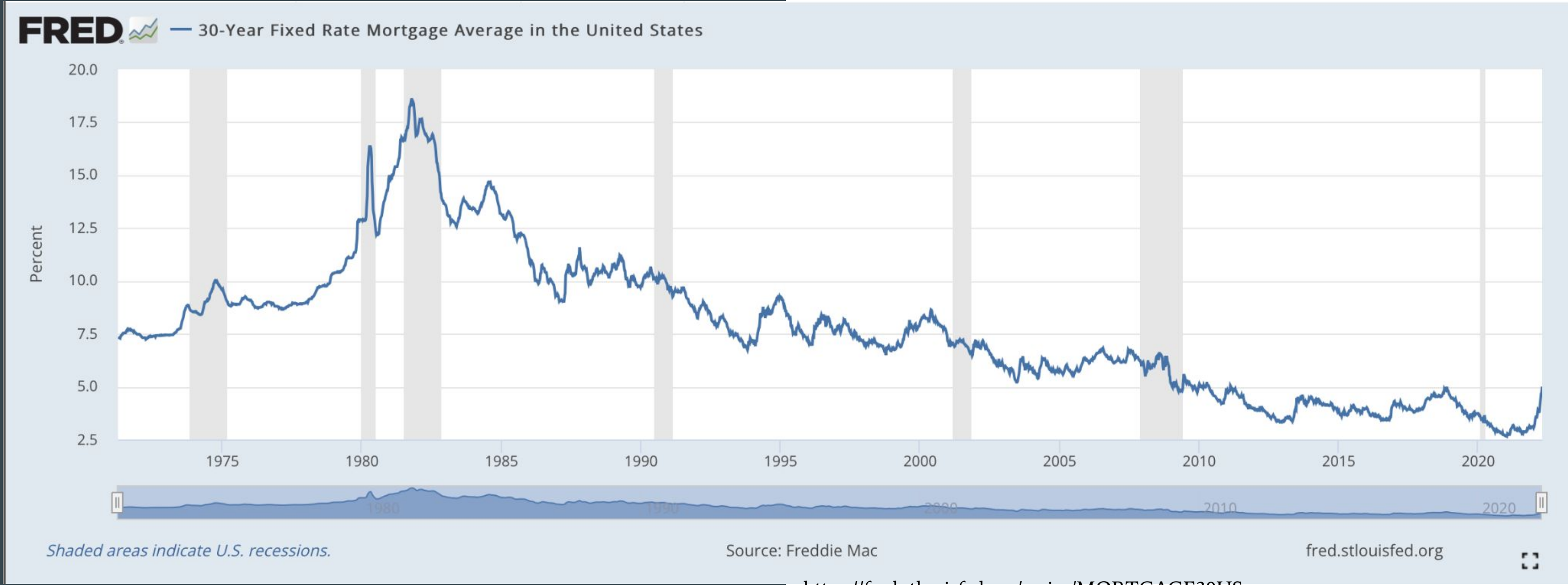
- Trends could change:
  - When/if you buy
  - When you refinance
  - Show the current economic stability

Overview of Data

30-Year Fixed Rate Mortgage Average in the United States

Data from: FRED Economic Data (U.S Federal Reserve)

- Data from 1971 to Present
- Updated every Thursday with the average rate
- Includes rates during times of recession



https://fred.stlouisfed.org/series/MORTGAGE30US

Original Research Questions:

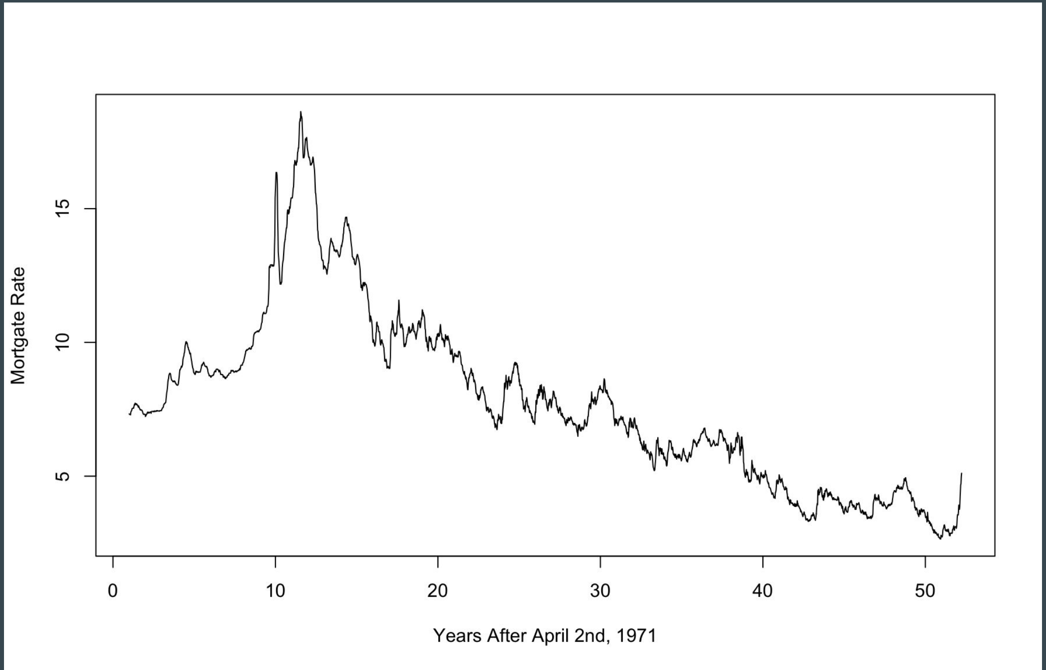
- Will COVID years affect the current rate?
- Does the current inflation affect the predicted rate?

- Looking at the data it seems to be unaffected by COVID at the moment. Seeing as inflation is just now starting to grow nationally, the mortgage rate is slower to react to that. Therefore there currently isn’t a spike in rates, but could be in the future.

New Research Questions:

- Can we form a model that accurately forecasts the average rate?
- What will it forecast for today’s rate?

Raw Data



The data does not appear to have a deterministic trend and again, it appears to not be affected by COVID (as COVID was in the last 3 years)

T-test and HAC Test

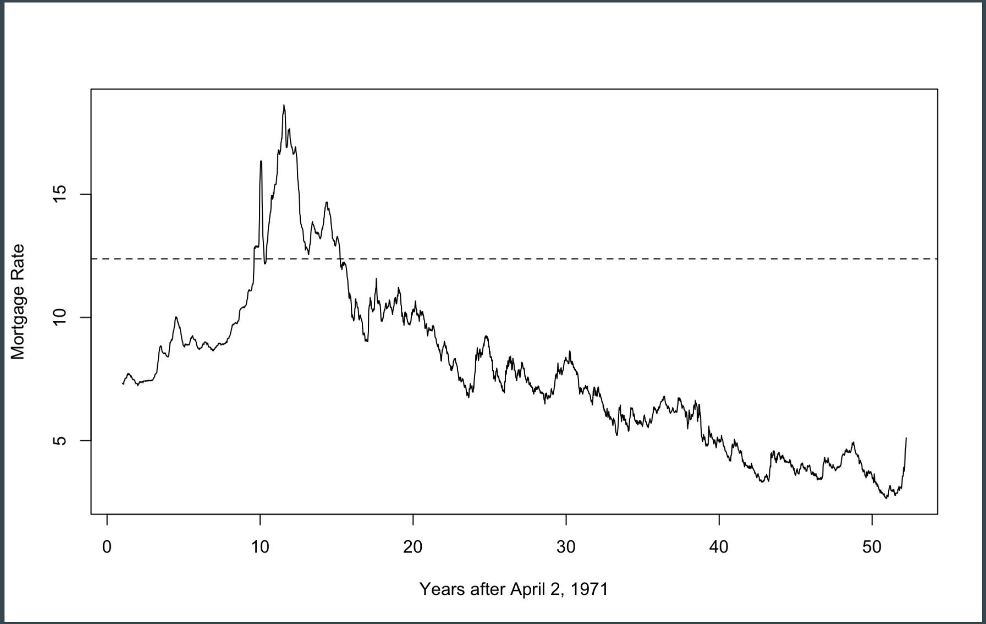
```
> summary(fm)

Call:
lm(formula = data ~ time(data))

Residuals:
    Min      1Q  Median      3Q     Max 
-4.9027 -0.8066 -0.2511  0.6879  8.2483 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 12.378749   0.082664  149.75   <2e-16 ***
time(data)  -0.172792   0.002715  -63.65   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.073 on 2663 degrees of freedom
Multiple R-squared:  0.6034,	Adjusted R-squared:  0.6032 
F-statistic:  4051 on 1 and 2663 DF,  p-value: < 2.2e-16
```

```
> coeftest(fm, vcov=vcovHAC(fm))

t test of coefficients:

             Estimate Std. Error t value  Pr(>|t|)    
(Intercept) 12.378749   1.295463  9.5555 < 2.2e-16 ***
time(data)  -0.172792   0.034613 -4.9920  6.36e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The t-test returned a R-squared value of 0.6034 and the HAC test shows the p-value is significant.

Data with the Fitted Constant Trend

We can see that the fitted constant trend is largely affected by the first 20 years of data.



Standard Errors

We can see the standard errors of the lm function and HAC estimation at right.

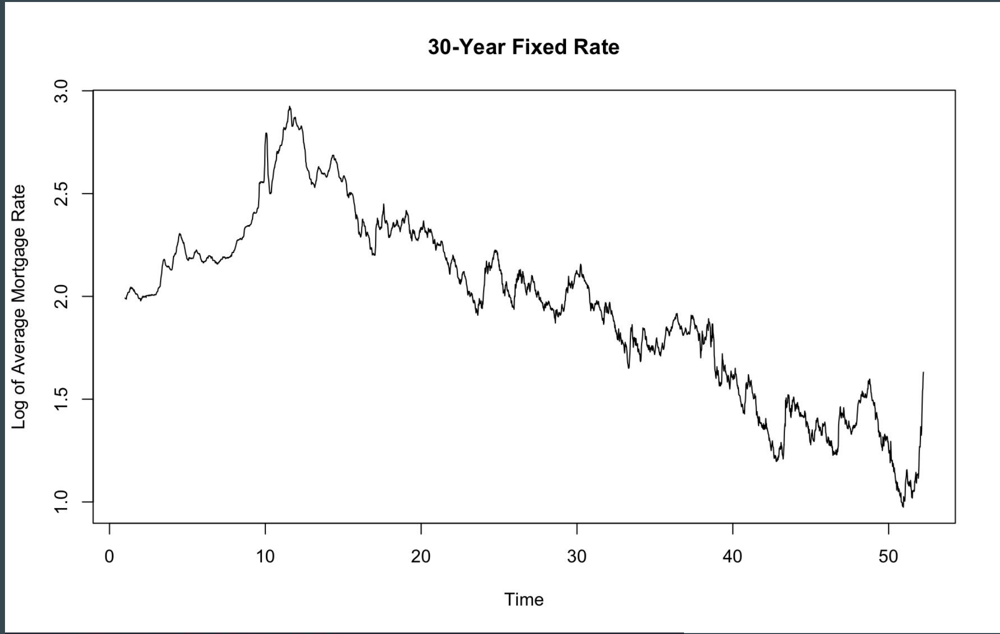
Their theoretical value being: 245.5919

```
> #######compare the standard errors
> ###standard error in 'lm' function
> se1=summary(fm)$coeff[1,2]
> se1
[1] 0.08266433
> ###standard error from HAC estimation
> se2=coeftest(fm, vcov=vcovHAC(fm))[1,2]
> se2
[1] 1.295463
```

Log Transformation

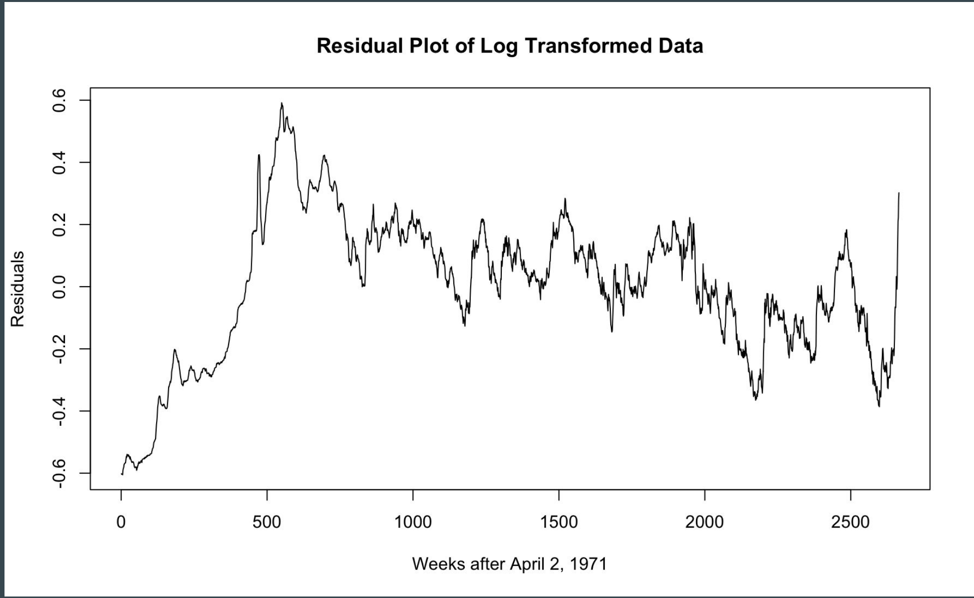
We then performed a log transformation which smoothed the data slightly, but did not affect the overall shape.

And the Shapiro Wilk's test for normality on the Log Transformed data did return with a significant p-value




```
> shapiro.test(ldata)

        Shapiro-Wilk normality test

data:  ldata
W = 0.98026, p-value < 2.2e-16
```

Residuals of Log Transformed Data




```
> summary(model)

Call:
lm(formula = ldata ~ c(1:length(data)))

Residuals:
     Min       1Q   Median       3Q      Max 
-0.60525 -0.13901  0.02104  0.15215  0.59172 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)        2.594e+00  8.917e-03  290.92   <2e-16 ***
c(1:length(data)) -4.747e-04  5.794e-06  -81.93   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2301 on 2663 degrees of freedom
Multiple R-squared:  0.7159,	Adjusted R-squared:  0.7158 
F-statistic:  6712 on 1 and 2663 DF,  p-value: < 2.2e-16
```

We can see that the R-squared value for the residuals is 0.7159

Tests for Log Transformed Data

Using an alpha value of 0.05, we can see that all the p-values for Phillips-Perron and Augmented Dickey-Fuller are not significant but the KPSS test is.

```
> pp.test(ldata)

        Phillips-Perron Unit Root Test

data:  ldata
Dickey-Fuller Z(alpha) = -12.07, Truncation lag parameter = 9, p-value = 0.4365
alternative hypothesis: stationary

> kpss.test(ldata)

        KPSS Test for Level Stationarity

data:  ldata
KPSS Level = 21.213, Truncation lag parameter = 9, p-value = 0.01

Warning message:
In kpss.test(ldata) : p-value smaller than printed p-value
> adf.test(ldata)

        Augmented Dickey-Fuller Test

data:  ldata
Dickey-Fuller = -2.819, Lag order = 13, p-value = 0.2316
alternative hypothesis: stationary
```

Finding a Model

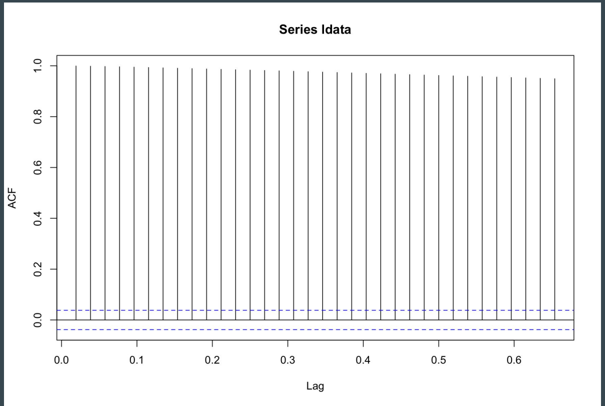
We ran a PCAF and an ACF plot for the log transformed data and we can see that they both suggest an ARIMA (1,1,0) model.

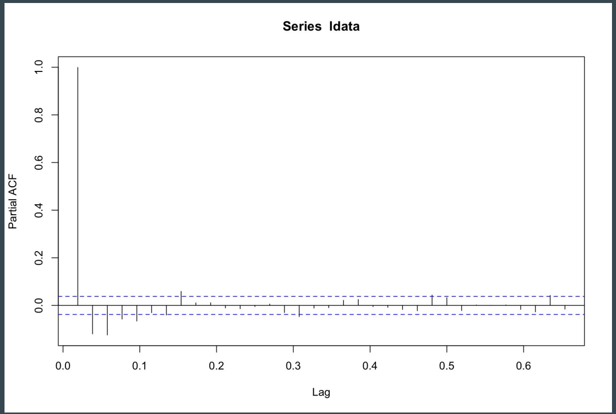
The Auto-ARIMA function suggests an ARIMA (1,1,2) and the eacf suggests ARIMA(2,1,1)

```
Series: ldata
ARIMA(1,1,2)

Coefficients:
         ar1      ma1     ma2
      0.6905  -0.5877  0.0484
s.e.  0.0619   0.0640  0.0230

sigma^2 = 0.0002122:  log likelihood = 7487.74
AIC=-14967.47   AICc=-14967.46   BIC=-14943.92
```

AR/MA

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	x	x	x	x	x	x	x	x	x	x	x	x	x
1	x	x	x	x	x	x	x	o	o	o	o	o	o	o
2	x	o	o	o	o	o	o	o	o	o	o	o	o	o
3	x	x	o	o	o	x	o	o	o	o	o	o	o	o
4	x	x	x	o	o	o	o	o	o	o	o	o	o	o
5	x	x	x	x	o	o	o	o	x	o	o	o	o	o
6	x	x	x	x	x	o	x	x	o	o	o	o	o	o
7	x	x	x	x	x	x	o	o	o	o	o	o	o	o





Models

Model 1: ARIMA(1,1,0)

AIC = -14917.22 AICc = -14917.21 BIC =-14899.55

Model 2: ARIMA(1,1,2)

AIC = -14965.56 AICc = -14965.53BIC = -14936.12

Model 3: ARIMA(2,1,1)

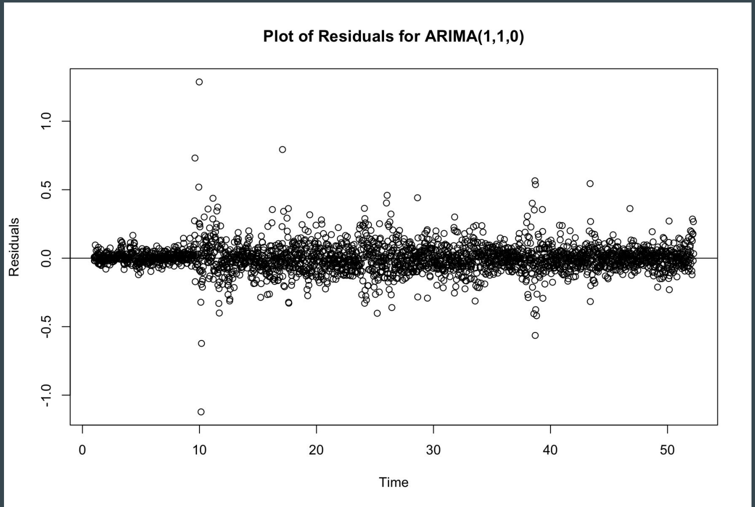
AIC = -14965.29 AICc = -14965.27BIC =-14935.86

We can see that model 1 returns the best AIC and AICc values.

Residual Analysis for ARIMA(1,1,0)

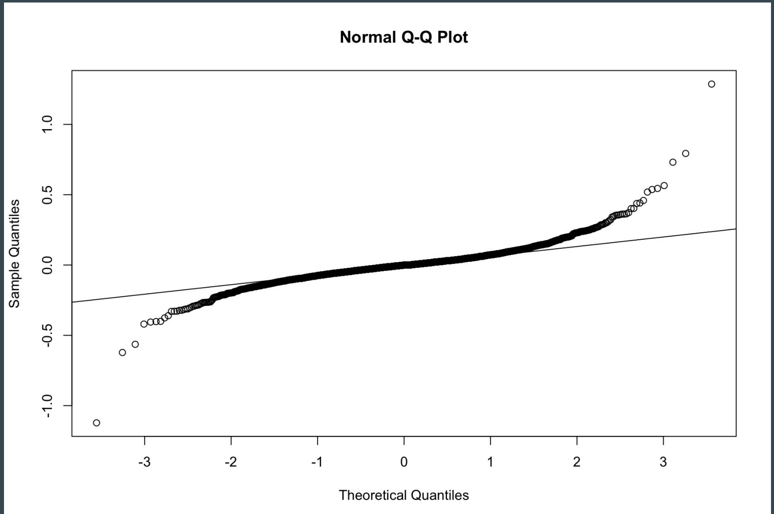
We can see from the plots to the right that the center of the data is normal, but has a visible skew on both ends.

With the Shapiro Wilks and Box-Ljung test returning significant p-values.

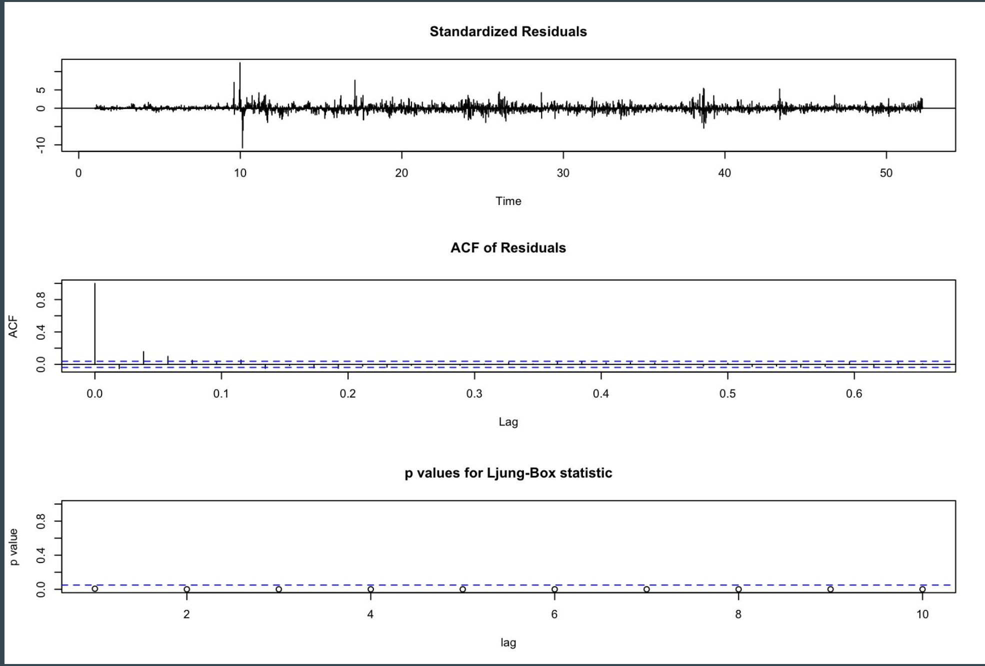



```
> shapiro.test(residualsm1)

        Shapiro-Wilk normality test

data:  residualsm1
W = 0.87526, p-value < 2.2e-16
```

```
        Box-Ljung test

data:  residualsm1
X-squared = 7.3697, df = 1, p-value = 0.006633
```



Forecasting the Model



Forecasting

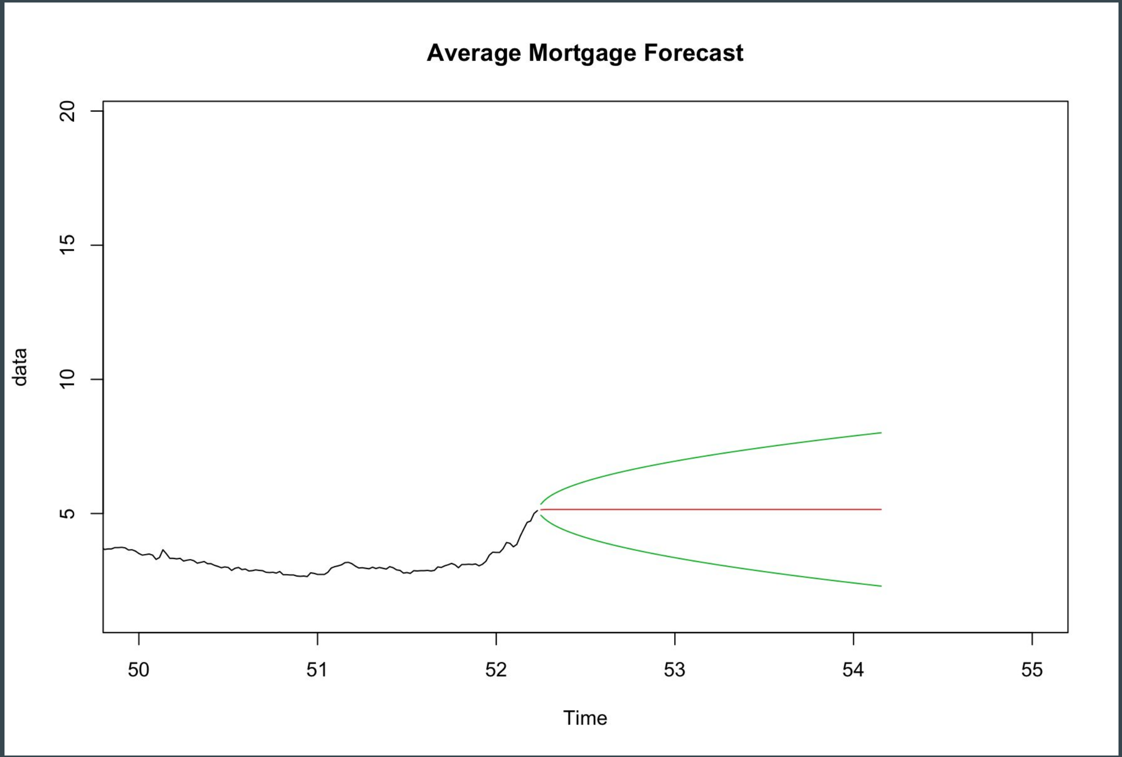
Data cuts off at 4/14/2022

Actual rate for 4/21/2022:

5.11

Predicted rate for 4/21/2022:

5.0936



Possible Improvement

- ❖ Choosing more detailed data
  - ➢ Daily rates
  - ➢ More factors to explain variability
- ❖ Building a more complex model