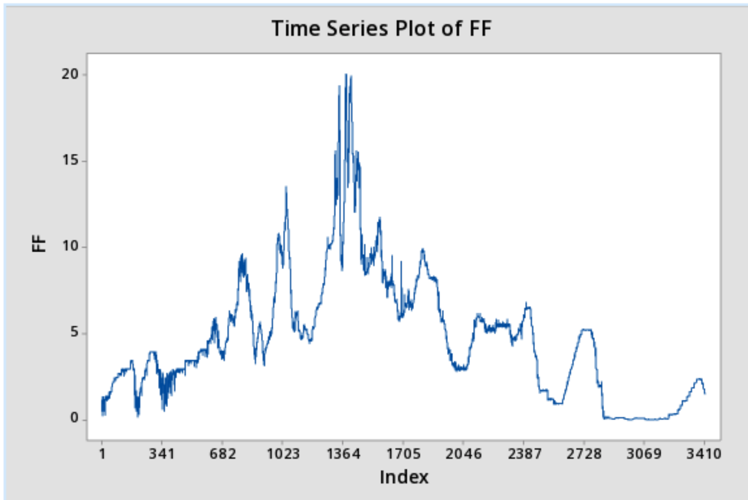
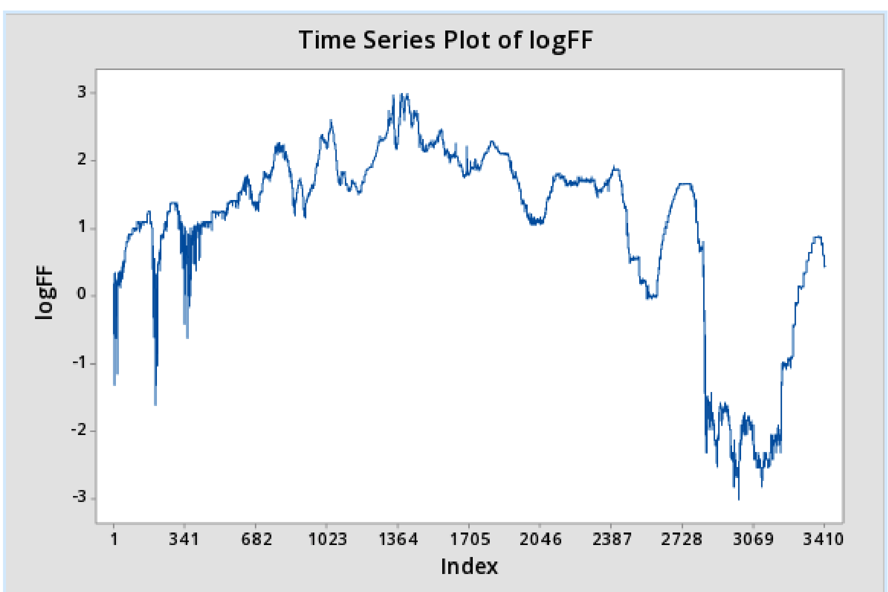
Trevor Mitchell
Times Series Forecasting – Stat 2302
Project 2

The data I have chosen to build an ARIMA-ARCH model for is the effective federal funds rate from 1954-07-07 to 2019-12-11. Here is the time series plot of the data:

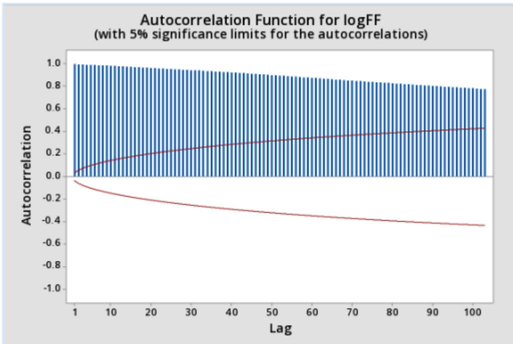


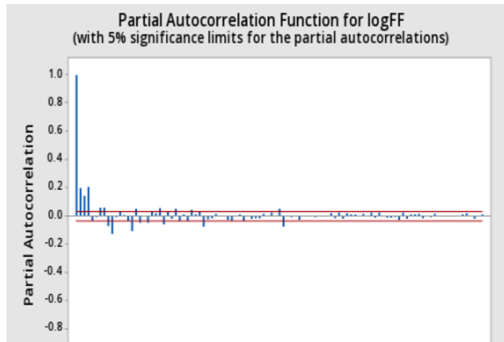
Observations and Model Building

In order to check for any level-dependent volatility in the dataset I decided to take logs. After taking logs I was able to remove a great deal of level-dependent volatility.



After taking logs, however, there did appear to still be a trend component in the data. The ACF of the original time series dies down very slowly, so it seemed very appropriate to difference the data.



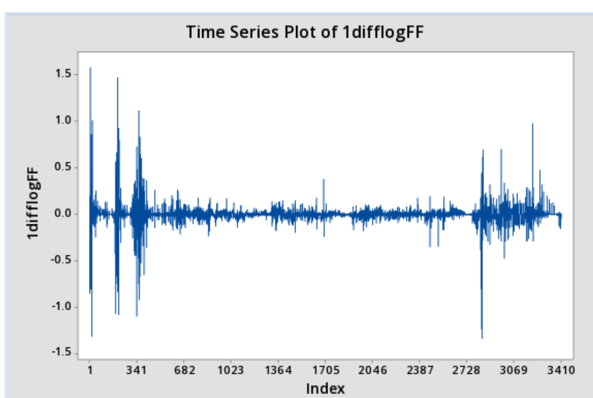


After differencing the data, the first lag of the ACF of the differenced data became negative. However, it was only -.201239 as opposed to the original .995763 of the log federal funds rate dataset. I also checked the 1st lag of the second difference which was -.533977, so I reasoned that differencing twice was too much. In addition, the first difference's ACF plot goes to zero much quicker than the dataset before differencing and yields the smallest standard deviation with value of 0.11897 compared to values of 1.3124 (Original Series) and 0.18440 (2nd differenced series).

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
logFF	3415	0	1.0611	0.0225	1.3124	-2.9957	0.7467	1.5019	1.8718
1difflogFF	3414	1	0.00013	0.00204	0.11897	-1.32687	-0.01704	0.00000	0.01914
2difflogFF	3413	2	-0.00006	0.00316	0.18440	-1.75815	-0.03651	0.00000	0.02935

The first difference also yields a time series that fluctuates around a well-defined mean. Differencing the series once appears to make it stationary.



After figuring out how much differencing my data needed, I used an AIC_C grid search to find the optimal parameters for my ARIMA model. I allowed AIC_C to determine whether or not a constant was needed in my model. As a result of my AIC_C grid search, I found ARIMA_414c (with constant) to be my optimal model. I used the respective AIC_C formula for models that do and do not include a constant term.

$$AIC_C = N \log\left(\frac{SS}{N}\right) + 2(p+q+1)\,\frac{N}{N-p-q-2} \quad \text{if no constant term in model ,}$$

$$AIC_C = N \log\left(\frac{SS}{N}\right) + 2(p+q+2)\,\frac{N}{N-p-q-3} \quad \text{if constant term is included .}$$

Model	AIC_C Score	SS	P	Q	N	REMARKS	BEST MODEL
ARIMA_010	-6309.023662	48.3084	0	0	3413		FALSE
ARIMA_011	-6404.950716	45.2199	0	1	3413		FALSE
ARIMA_012	-6444.548032	43.9684	0	2	3413		FALSE
ARIMA_013	-6452.016172	43.6883	0	3	3413		FALSE
ARIMA_014	-6475.02479	42.9572	0	4	3413		FALSE
ARIMA_015	-6480.971789	42.7273	0	5	3413		FALSE
ARIMA_110	-6368.388459	46.3492	1	0	3413		FALSE
ARIMA_111	-6442.284339	44.0356	1	1	3413		FALSE
ARIMA_112	-6445.330536	43.8858	1	2	3413		FALSE
ARIMA_113	-6461.006265	43.3654	1	3	3413		FALSE
ARIMA_114	0	NA	1	4	3413		FALSE
ARIMA_115	0	NA	1	5	3413		FALSE
ARIMA_210	-6406.872715	45.1003	2	0	3413		FALSE
ARIMA_211	0	NA	2	1	3413		FALSE
ARIMA_212	-6462.045714	43.335	2	2	3413		FALSE
ARIMA_213	-6504.300153	42.0601	2	3	3413		FALSE
ARIMA_214	-6503.391855	42.0289	2	4	3413		FALSE
ARIMA_215	-6492.837163	42.2719	2	5	3413		FALSE
ARIMA_310	-6480.824185	42.8474	3	0	3413		FALSE
ARIMA_311	-6478.821771	42.8473	3	1	3413		FALSE
ARIMA_312	-6490.29473	42.4594	3	2	3413		FALSE
ARIMA_313	-6503.62111	42.0224	3	3	3413		FALSE
ARIMA_314	-6502.218516	42.0052	3	4	3413		FALSE
ARIMA_315	0	NA	3	5	3413		FALSE
ARIMA_410	-6478.825231	42.8472	4	0	3413		FALSE
ARIMA_411	-6477.195299	42.8363	4	1	3413		FALSE
ARIMA_412	-6493.682885	42.3051	4	2	3413		FALSE
ARIMA_413	0	NA	4	3	3413		FALSE

ARIMA_414	0	NA	4	4	3413		FALSE
ARIMA_415	-6467.847893	42.8741	4	5	3413		FALSE
ARIMA_510	-6478.136789	42.8091	5	0	3413		FALSE
ARIMA_511	-6476.654945	42.7939	5	1	3413		FALSE
ARIMA_512	-6505.150253	41.9222	5	2	3413		FALSE
ARIMA_513	0	NA	5	3	3413		FALSE
ARIMA_514	0	NA	5	4	3413		FALSE
ARIMA_515	0	NA	5	5	3413		FALSE
ARIMA_010_c	-6293.633595	48.7467	0	0	3413		FALSE
ARIMA_011_c	-6402.953751	45.2197	0	1	3413		FALSE
ARIMA_012_c	-6442.553449	43.9681	0	2	3413		FALSE
ARIMA_013_c	-6450.020477	43.688	0	3	3413		FALSE
ARIMA_014_c	-6473.02809	42.9569	0	4	3413		FALSE
ARIMA_015_c	-6478.973966	42.727	0	5	3413		FALSE
ARIMA_110_c	-6366.384938	46.3492	1	0	3413		FALSE
ARIMA_111_c	-6440.28974	44.0353	1	1	3413		FALSE
ARIMA_112_c	-6443.334795	43.8855	1	2	3413		FALSE
ARIMA_113_c	-6459.006049	43.3652	1	3	3413		FALSE
ARIMA_114_c	-6479.369497	42.7156	1	4	3413		FALSE
ARIMA_115_c	-6476.961086	42.7271	1	5	3413		FALSE
ARIMA_210_c	-6404.871304	45.1002	2	0	3413		FALSE
ARIMA_211_c	0	NA	2	1	3413		FALSE
ARIMA_212_c	-6460.042083	43.3349	2	2	3413		FALSE
ARIMA_213_c	-6502.538631	42.0531	2	3	3413		FALSE
ARIMA_214_c	-6501.393024	42.0286	2	4	3413		FALSE
ARIMA_215_c	0	NA	2	5	3413		FALSE
ARIMA_310_c	-6478.82869	42.8471	3	0	3413		FALSE
ARIMA_311_c	-6476.825098	42.847	3	1	3413		FALSE
ARIMA_312_c	-6488.2865	42.4594	3	2	[illegible]3		FALSE
ARIMA_313_c	-6501.657555	42.0211	3	3	3413		FALSE
ARIMA_314_c	-6499.361273	42.0292	3	4	3413		FALSE
ARIMA_315_c	0	NA	3	5	3413		FALSE
ARIMA_410_c	-6476.825098	42.847	4	0	3413		FALSE
ARIMA_411_c	-6475.19745	42.836	4	1	3413		FALSE
ARIMA_412_c	-6491.747055	42.303	4	2	3413		FALSE
ARIMA_413_c	0	NA	4	3	3413		FALSE
ARIMA_414_c	-6522.854167	41.3122	4	4	3413		TRUE
ARIMA_415_c	-6471.19985	42.7192	4	5	3413		FALSE
ARIMA_510_c	-6476.138946	42.8088	5	0	3413		FALSE
ARIMA_511_c	-6474.655926	42.7936	5	1	3413		FALSE
ARIMA_512_c	-6503.121983	41.9227	5	2	3413		FALSE
ARIMA_513_c	0	NA	5	3	3413		FALSE
ARIMA_514_c	0	NA	5	4	3413		FALSE
ARIMA_515_c	0	NA	5	5	3413		FALSE

Complete form of Fitted Model (ARMA)

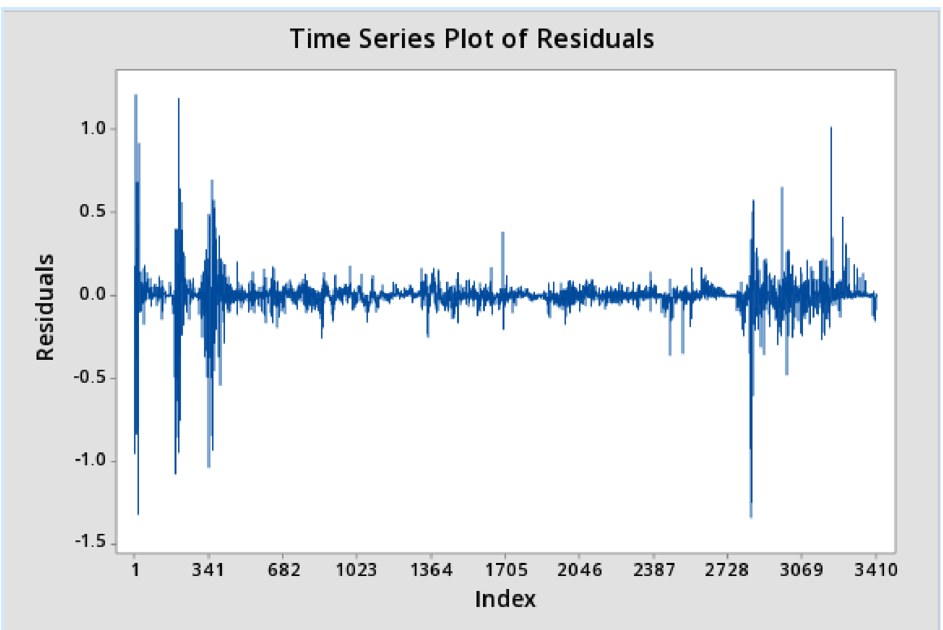
$$x_t = .00039 - 0.4899x_{t-1} - 0.4732x_{t-2} - 0.7069x_{t-3} + 0.2548x_{t-4} + 0.2331\epsilon_{t-1} + 0.1971\epsilon_{t-2} + 0.4316\epsilon_{t-3} - 0.4509\epsilon_{t-4}$$

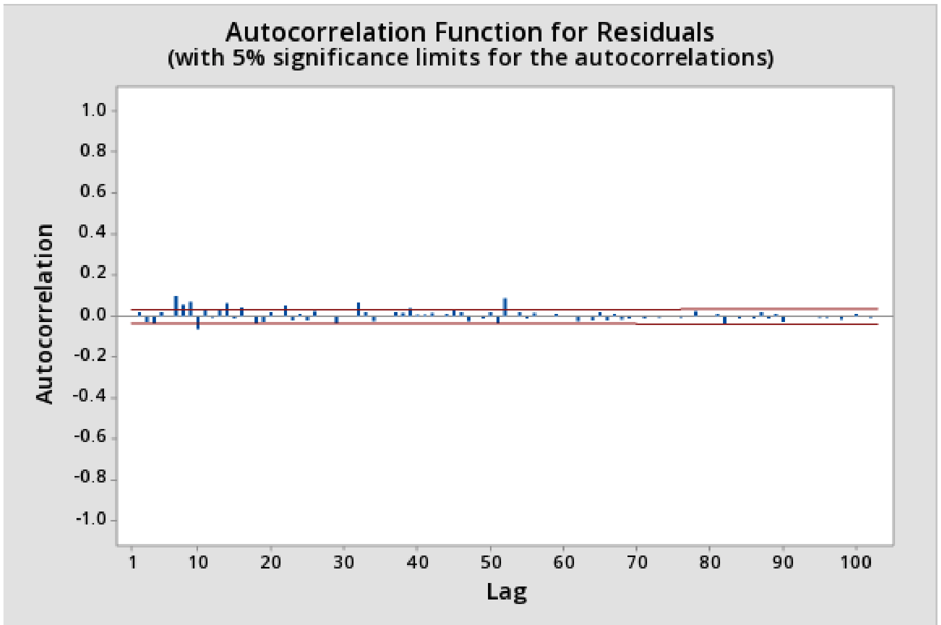
Ljung-Box Statistics

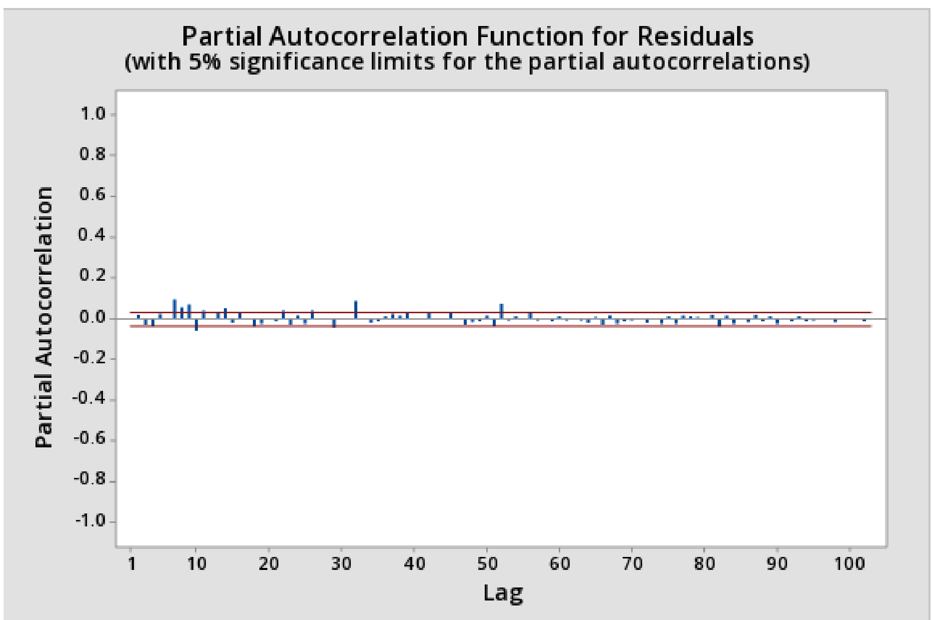
The Ljung Box statistics seemed to indicate that my ARIMA model is not adequate, as my p-values appear to be zero which is much less .05. However, the ACF and PACF plot of the residuals seem to indicate that my model chosen is okay. I have autocorrelations and partial autocorrelations near zero for most lags. Most of the lags that are a bit farther from zero are not statistically significant and those that are significant are barely so. The plot itself also appears to have zero mean.

Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

Lag	12	24	36	48
Chi-Square	86.71	131.29	158.24	174.16
DF	3	15	27	39
P-Value	0.000	0.000	0.000	0.000

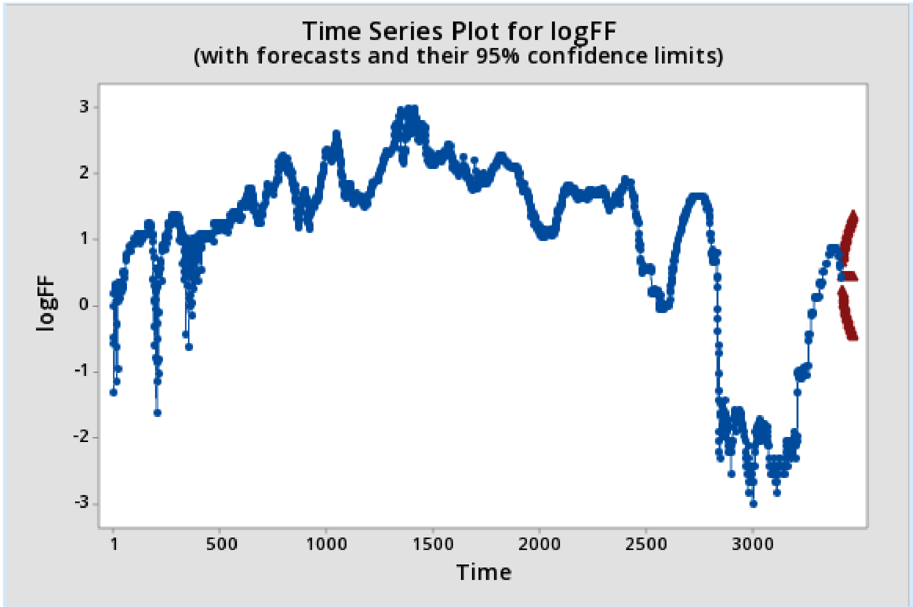






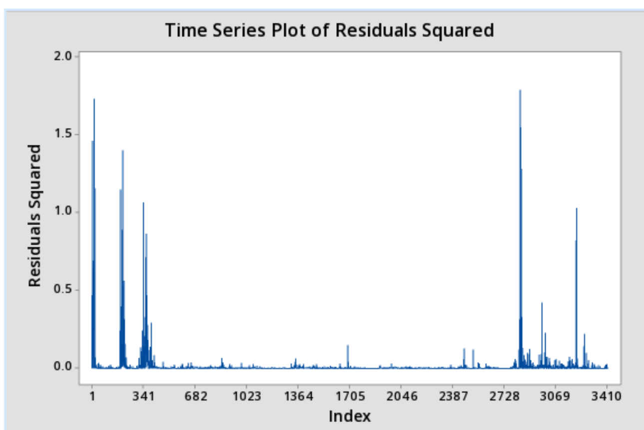
Forecasts of ARIMA_012 at Lead Times 1-50

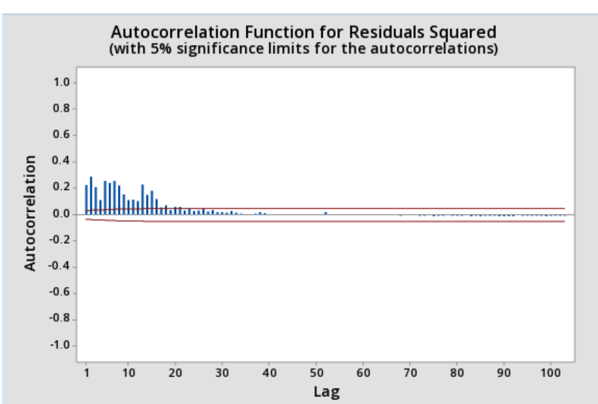
The forecast interval appears to be a bit too narrow. Much of the historical datapoints fall outside of the forecasts. However, recent data points do appear to be within the forecast interval.

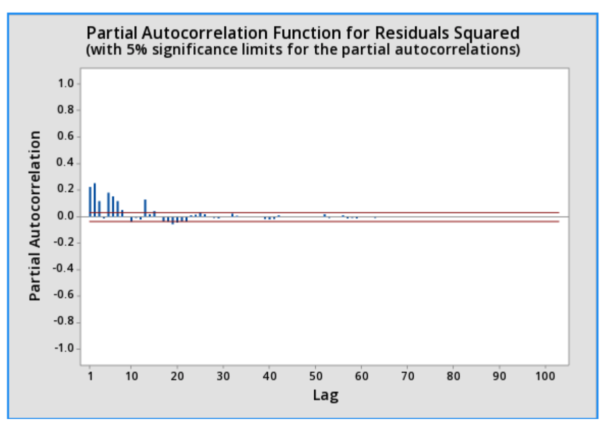


ARCH Component

Plots of Residuals Squared Time Series, ACF, & PACF







While the residuals are approximately uncorrelated, the ACF and PACF of the squared residuals have significant lags which is reasonable evidence that the residuals are not independent and instead suggest evidence of conditional heteroscedasticity.

According to AIC_C the optimal ARCH(q) model is the ARCH(10). The parameter values for a0, a1, a2, a4, a5, a6,a7, and a10 are quite statistically significant. While a9 is significant at .01 level of significance, a8 is significant at .05, and a3 is not statistically significant. The complete form of the ARCH(10) model is as follows below:

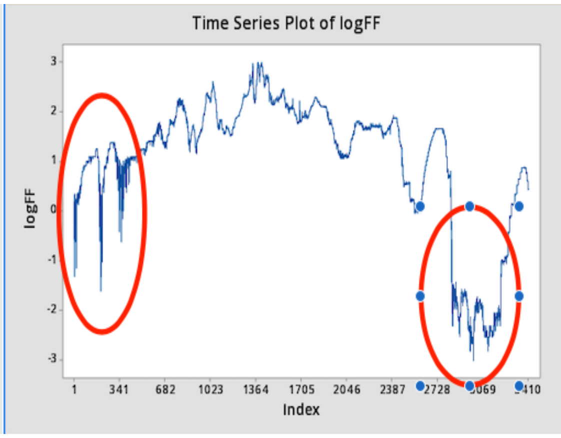
$$h_t = 5.564e-04 + 3.979e01\epsilon_{t-1}^2 + 3.490e01\epsilon_{t-2}^2 + 6.163e03\epsilon_{t-3}^2 + 7.012e02\epsilon_{t-4}^2 + 3.302e02\epsilon_{t-5}^2 + 9.630e02\epsilon_{t-6}^2 + 9.742e02\epsilon_{t-7}^2 + 1.390e02\epsilon_{t-8}^2 + 1.485e02\epsilon_{t-9}^2 + 1.767e01\epsilon_{t-10}^2$$

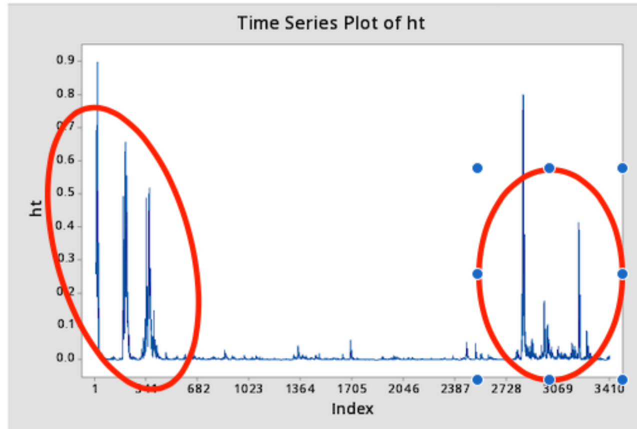
Model	AIC_C Score	logLik	Q	N	REMARKS	BEST MODEL
ARCH_0	-5377.9508	2689.976	0	3413		FALSE
ARCH_1	-7357.9765	3680.99	1	3413		FALSE
ARCH_2	-9434.897	4720.452	2	3413		FALSE
ARCH_3	-9612.0703	4810.041	3	3413		FALSE
ARCH_4	-9722.4884	4866.253	4	3413		FALSE
ARCH_5	-9756.8893	4884.457	5	3413		FALSE
ARCH_6	-9869.7991	4941.916	6	3413		FALSE
ARCH_7	-9889.8997	4952.971	7	3413		FALSE
ARCH_8	-9894.1911	4956.122	8	3413		FALSE
ARCH_9	-9893.9253	4956.995	9	3413		FALSE
ARCH_10	-10008.296	5015.187	10	3413		TRUE
GARCH_11	-9907.257	4956.632	2	3413		FALSE

The unconditional variance of the shocks was computed to be 0.014975856230972646 by taking the expectation.

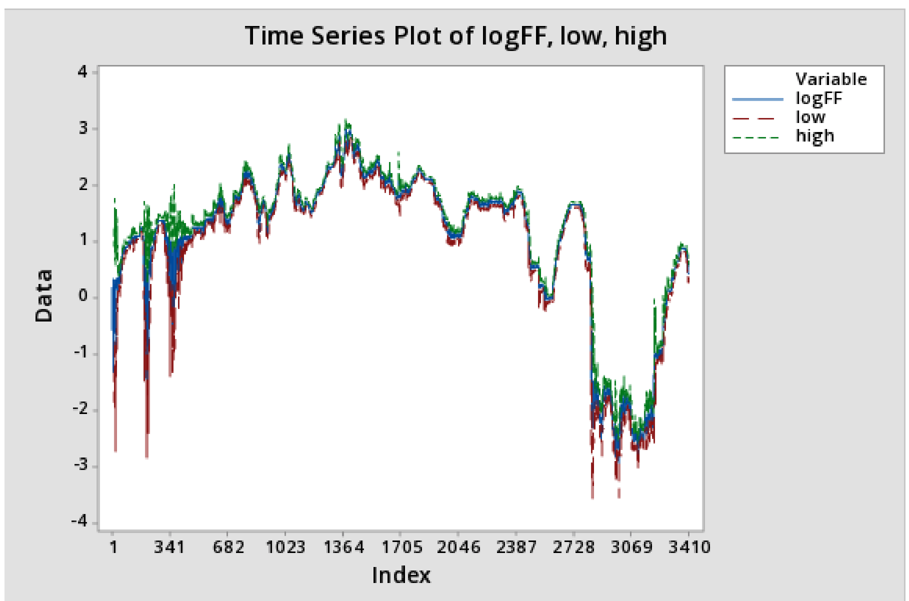
The 95% one-step ahead forecast interval for the ARIMA_ARCH model is (0.281189, 0.608628). The 5th percentile of the conditional distribution is 0.30708375783198616 for the next periods log federal funds rate.

As one can see from the plots below, the highly volatile periods for the log federal funds rate appear to coincide with the highly volatile periods of the conditional variances plot

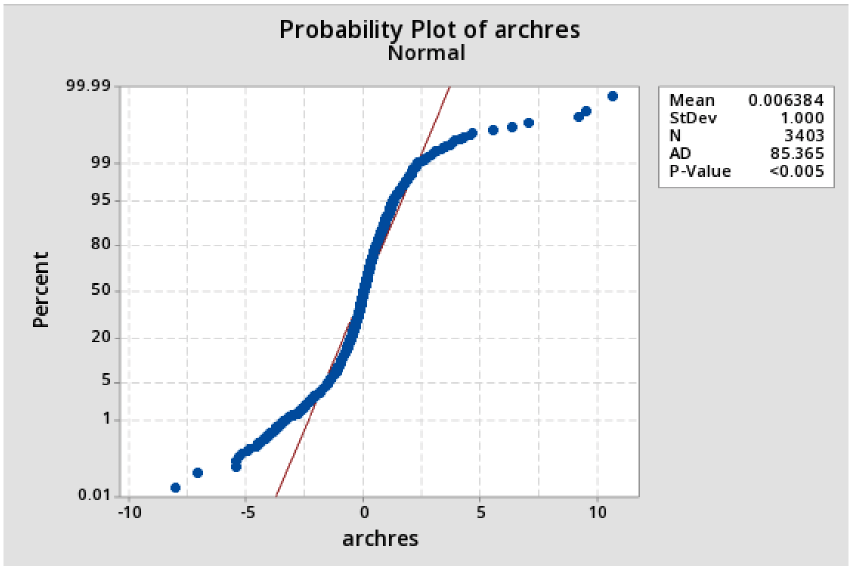




The forecast intervals appear to be very useful and practical as the log federal funds rate appears to fall relatively tightly in between the upper and lower bounds of the ARIMA-ARCH intervals.



In accordance with the normality tests of my ARIMA-ARCH residuals, my model does not seem to follow a normal distribution as I have a p-value less than .005. The model does not seem to adequately describe “long-tailedness” in the data. The model appears to show increasing departure from the fitted line above and below for data points on the end.



The ARIMA-ARCH intervals failed 169 times which accounted for 4.95021 % of the time.

The one-step ahead forecast for the ARIMA model and the ARIMA-ARCH model both contained the actual target value of 0.438255 (the final omitted data point) with ARIMA having an interval of (0.228941, 0.66088) and ARIMA-ARCH having an interval (0.281189, 0.608628). While both intervals worked, the ARIMA-ARCH model had a tighter upper and lower bound on the target. It would seem that the ARIMA-ARCH generated a superior prediction interval.

Source of the Data Set

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