Shafim Khoda

FIN 550

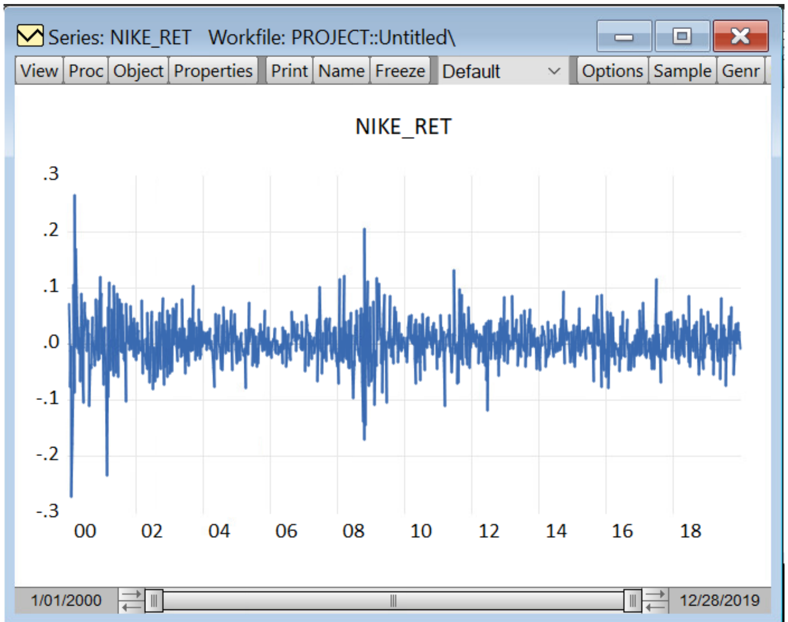
Nike Time Series Project

Executive Summary:

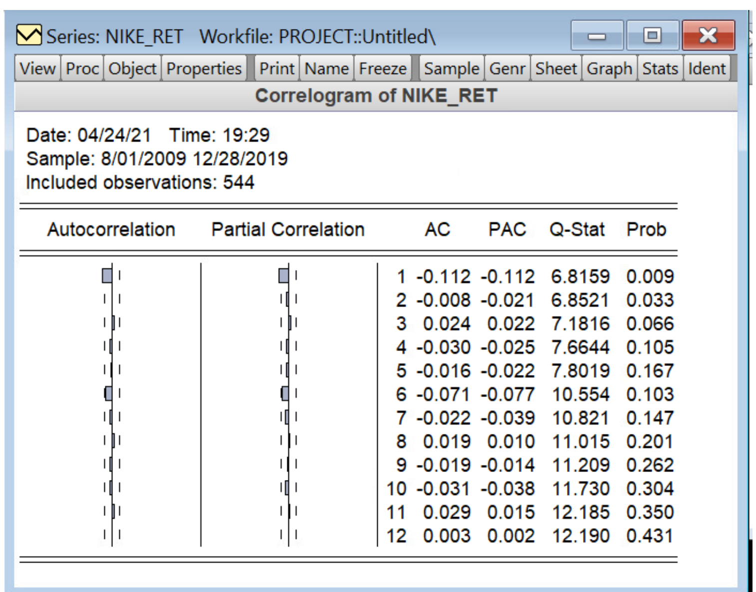
This report summarizes the Nike stock returns from the time of 01/01/2000 to 12/31/2019. For this report we looked at weekly returns. We first evaluated if the returns had any dependence then modeled the conditional mean with an ARMA model and then proceeded to model the conditional variance with a GARCH process. To check the validity of these models we looked at the AIC, p-values, residuals, residuals squared, and even used the principle of parsimony. Using all the principles we were able to select the best model for the Nike weekly returns. After this we did an in sample and out of sample value at risk analysis. We used a baseline set of values for our theoretical portfolio and proceeded to look at the in-sample VaR and looked at the out of sample VaR using the rolling window approach. Once this is finished, we assessed the performance of the portfolio and if the in-sample and out-of-sample measurements are comparable.

Data Description:

The data used for this analysis is from Yahoo Finance. Once I downloaded the closing price for Nike, I changed the series into returns so that I would have a stationary dataset. This allows one to do time series analysis.



As you can see here, all the following criteria of stationarity seems to be met so a standard time series analysis can commence. From here we look at the correlogram to see if serial correlation exists in this plot. Using the first 12 lags the correlogram we obtain looks like this:


Series: NIKE_RET Workfile: PROJECT::Untitled\

View Proc Object Properties Print Name Freeze Sample Genr Sheet Graph Stats Ident

Correlogram of NIKE_RET

Date: 04/24/21 Time: 19:29
Sample: 8/01/2009 12/28/2019
Included observations: 544

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.112	-0.112	6.8159	0.009
		2	-0.008	-0.021	6.8521	0.033
		3	0.024	0.022	7.1816	0.066
		4	-0.030	-0.025	7.6644	0.105
		5	-0.016	-0.022	7.8019	0.167
		6	-0.071	-0.077	10.554	0.103
		7	-0.022	-0.039	10.821	0.147
		8	0.019	0.010	11.015	0.201
		9	-0.019	-0.014	11.209	0.262
		10	-0.031	-0.038	11.730	0.304
		11	0.029	0.015	12.185	0.350
		12	0.003	0.002	12.190	0.431

Model Selection:

There is simply nothing special about this correlogram. It appears the mean represents a random walk and can be described as such. One would go with a random walk here due to the principle of parsimony. Overfitting a ARMA model of any sorts with a correlogram like this would provide minimal to no benefit however we will use an ARMA model here for practice. From here we essentially test different models based on geometric decays in the ACF and PACF. Since there is nothing much to be found here, I just took into consideration 1 step lags of MA and AR models to see which had the lowest AIC and the least amount of dependence in the correlogram. From here we ended up fitting an MA(2) model.

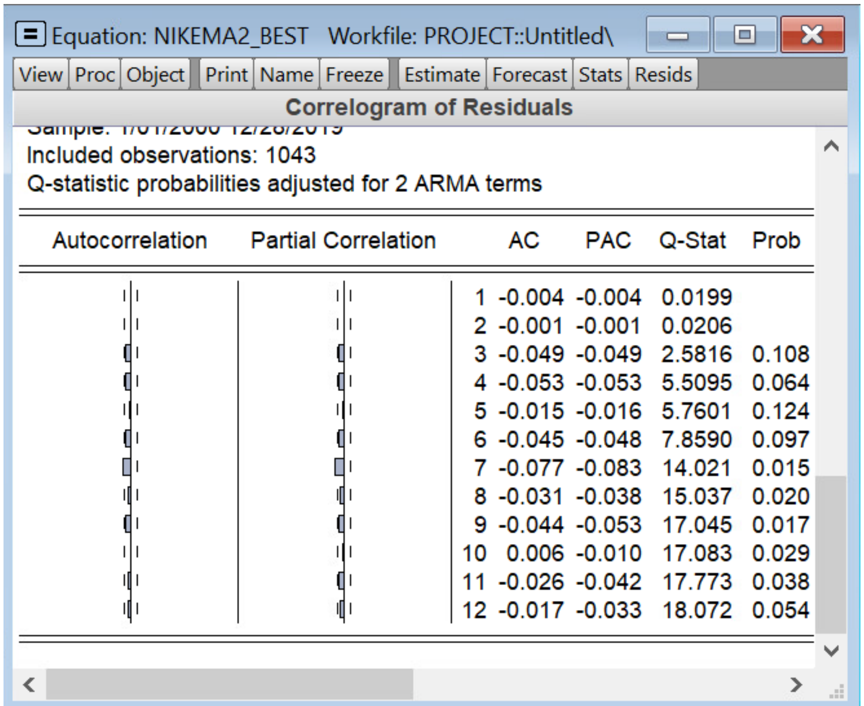
Equation: NIKEMA2_BEST Workfile: PROJECT::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: NIKE_RET
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 04/24/21 Time: 18:50
Sample: 1/08/2000 12/28/2019
Included observations: 1043
Convergence achieved after 40 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003507	0.001291	2.716968	0.0067
MA(1)	-0.072895	0.019912	-3.660860	0.0003
MA(2)	0.088514	0.020017	4.421941	0.0000
SIGMASQ	0.001557	3.76E-05	41.36639	0.0000

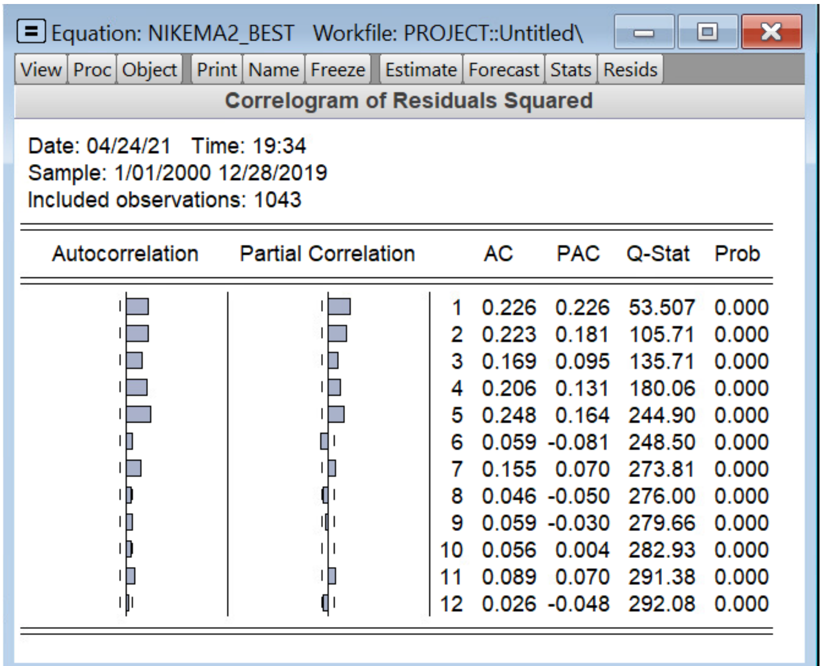
R-squared	0.013564	Mean dependent var	0.003507
Adjusted R-squared	0.010715	S.D. dependent var	0.039744
S.E. of regression	0.039530	Akaike info criterion	-3.619663
Sum squared resid	1.623570	Schwarz criterion	-3.600679
Log likelihood	1891.654	Hannan-Quinn criter.	-3.612462
F-statistic	4.762124	Durbin-Watson stat	2.005814



Sample: 1/01/2000 12/28/2019
Included observations: 1043
Q-statistic probabilities adjusted for 2 ARMA terms

	AC	PAC	Q-Stat	Prob
1	-0.004	-0.004	0.0199	
2	-0.001	-0.001	0.0206	
3	-0.049	-0.049	2.5816	0.108
4	-0.053	-0.053	5.5095	0.064
5	-0.015	-0.016	5.7601	0.124
6	-0.045	-0.048	7.8590	0.097
7	-0.077	-0.083	14.021	0.015
8	-0.031	-0.038	15.037	0.020
9	-0.044	-0.053	17.045	0.017
10	0.006	-0.010	17.083	0.029
11	-0.026	-0.042	17.773	0.038
12	-0.017	-0.033	18.072	0.054

For the next section we look at the volatility clustering in returns. Usually stocks returns are notorious for having some sort of inconsistent volatility that can be modeled via ARCH or GARCH models. After modeling the ARMA model we can see that there exists a decent amount of dependence within the squared residuals.


Equation: NIKEMA2_BEST Workfile: PROJECT::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Correlogram of Residuals Squared

Date: 04/24/21 Time: 19:34
Sample: 1/01/2000 12/28/2019
Included observations: 1043

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.226	0.226	53.507	0.000
		2	0.223	0.181	105.71	0.000
		3	0.169	0.095	135.71	0.000
		4	0.206	0.131	180.06	0.000
		5	0.248	0.164	244.90	0.000
		6	0.059	-0.081	248.50	0.000
		7	0.155	0.070	273.81	0.000
		8	0.046	-0.050	276.00	0.000
		9	0.059	-0.030	279.66	0.000
		10	0.056	0.004	282.93	0.000
		11	0.089	0.070	291.38	0.000
		12	0.026	-0.048	292.08	0.000

With this information we can now look ahead to several models to model the volatility of this stock return. Overall Nike does have many models that decreases the squared residual dependence of the returns however in a practical sense one must use the principle of parsimony to make sure that the model used is feasible and simple enough to use in practice. Given the low AIC, I think that the GARCH(1,1) is sufficient to model the volatility of Nike's stock. (Any model of higher order creates parameters with extremely high p-values.)

Equation: NIKEGARCH11_BEST Workfile: PROJECT::Untit...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Presample variance: backcast (parameter = 0.7)
GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.003609	0.001501	2.404829	0.0162
MA(1)	-0.090872	0.049688	-1.828839	0.0674
MA(2)	0.056030	0.043644	1.283799	0.1992
Variance Equation				
C	2.66E-05	9.56E-06	2.785002	0.0054
RESID(-1)^2	0.073941	0.015826	4.672117	0.0000
GARCH(-1)	0.908738	0.016338	55.62229	0.0000

R-squared	0.014385	Mean dependent var	0.002879
Adjusted R-squared	0.010411	S.D. dependent var	0.047702
S.E. of regression	0.047453	Akaike info criterion	-3.578262
Sum squared resid	1.116877	Schwarz criterion	-3.527609
Log likelihood	898.7764	Hannan-Quinn criter.	-3.558384
Durbin-Watson stat	1.937831		

Equation: NIKEGARCH11_BEST Workfile: PROJECT::Untit...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Correlogram of Standardized Residuals Squared

Date: 04/24/21 Time: 19:37
Sample: 1/01/2000 7/25/2009
Included observations: 499

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	-0.000	-0.000	0.0001	0.992
		2	0.050	0.050	1.2344	0.539
		3	0.035	0.035	1.8618	0.602
		4	0.007	0.004	1.8853	0.757
		5	-0.003	-0.006	1.8893	0.864
		6	-0.041	-0.043	2.7470	0.840
		7	0.021	0.021	2.9642	0.888
		8	-0.069	-0.064	5.3536	0.719
		9	0.012	0.013	5.4256	0.796
		10	-0.039	-0.034	6.2125	0.797
		11	0.029	0.032	6.6351	0.828
		12	-0.026	-0.025	6.9823	0.859

In conclusion of the model selection section we chose to model the conditional mean with a MA(2) and the conditional variance with a GARCH(1,1) model.

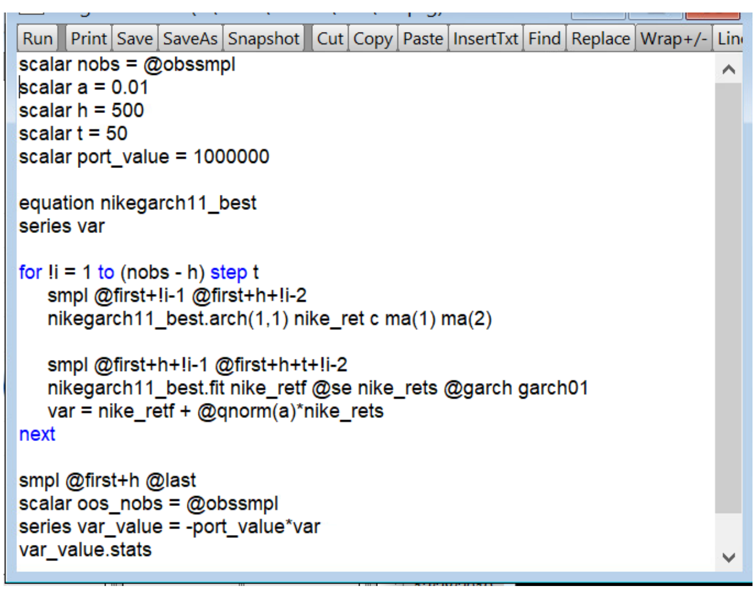
Value at Risk Analysis:

For the Value at Risk analysis we started with base set of conditions.

1. Alpha = 0.01
2. H = 500; Sample Period
3. T = 50; Step
4. Portfolio Value: $1,000,000

Side note: One complication I ran into when I first did this analysis was, I used daily data. Problem with this is that when one has that many data points, the underlying best model will change. Limiting the number of data points helps avoid. this issue.

So using the given inputs and weekly data, I then ran some code in order to find the in sample and out of sample value at risk.



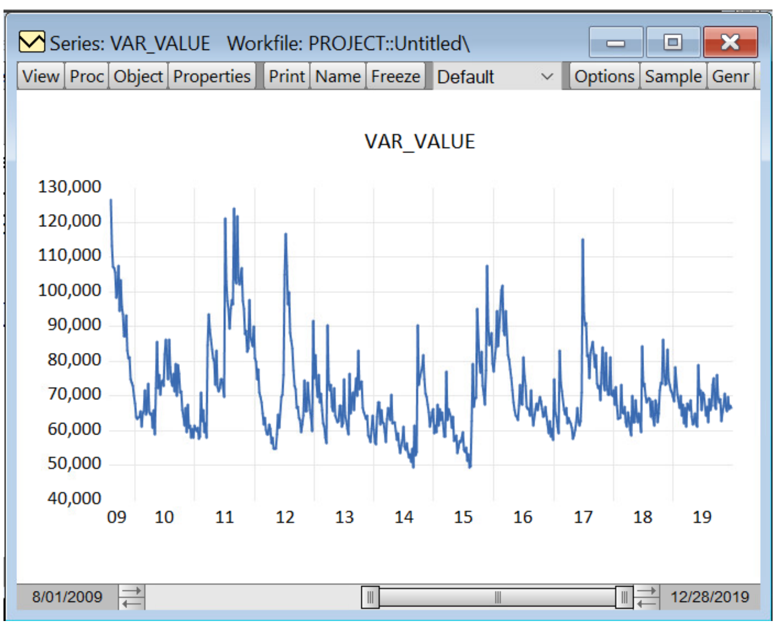

```
scalar nobs = @obssmpl
scalar a = 0.01
scalar h = 500
scalar t = 50
scalar port_value = 1000000

equation nikegarch11_best
series var

for !i = 1 to (nobs - h) step t
    smpl @first+!i-1 @first+h+!i-2
    nikegarch11_best.arch(1,1) nike_ret c ma(1) ma(2)

    smpl @first+h+!i-1 @first+h+t+!i-2
    nikegarch11_best.fit nike_retf @se nike_rets @garch garch01
    var = nike_retf + @qnorm(a)*nike_rets
next

smpl @first+h @last
scalar oos_nobs = @obssmpl
series var_value = -port_value*var
var_value.stats
```

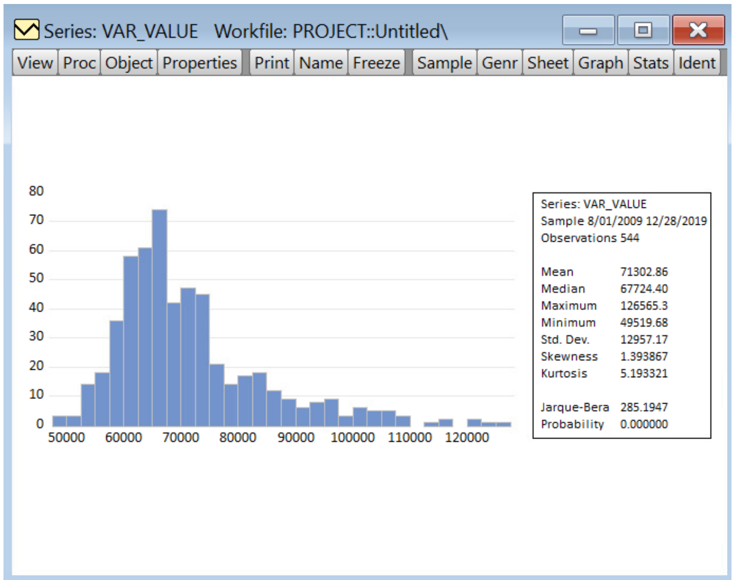
Of course, the first several lines is just defining some scalar quantities for our inputs for the computation later in the code. After this we define our most prominent volatility function

(GARCH(1,1)) and a new series called ‘var’ to represent the value at risk. After this we get into the for-loop which allows us to calculate the var function for in-sample and out-of-sample.

The first ‘smpl’ command sets up the window to estimate the chosen model while the second ‘smpl’ command sets up the out-of-sample period, that is, the estimated model is used to conduct static forecast for the weeks in the out-of-sample period. From here the for loop keeps shifting until we reach the end of the sample. After this the last section of code basically just calculates of the out of sample value at risk and stores it as series and presents us some basic statistics of the following value at risk.

So now let’s assess the results. With alpha = 0.01 and our portfolio value = $1,000,000 we can see what the 1% value at risk is. (There is a 1% chance that the portfolio drops by more than $x in the next week.)

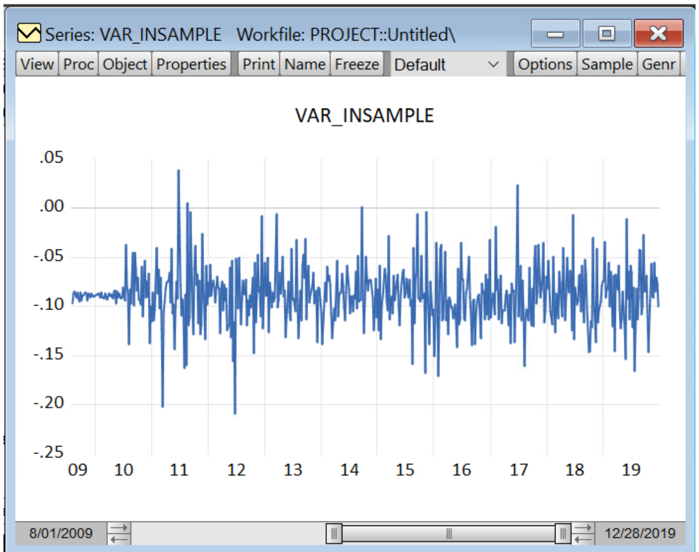


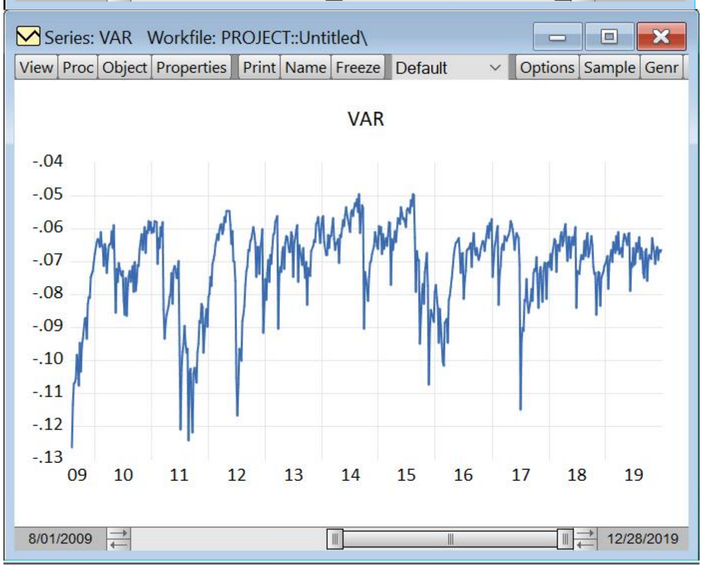


This is the out of sample value at risk analysis. This is the plot and histogram showing the values of what the value at risk would be given the returns from the period of 08/01/2009-12/28/2019. The mean signifies that on average, there is a 1% chance that portfolio loses more than $71,302.86 in a 1-week period.

This value at risk is good but it also depends if the goal of the portfolio is to be minimal risk. If the portfolio holder wants to have higher returns, over a certain period they most likely would need to combine this portfolio with a riskier asset to increase the value at risk.

For the next part we will look at how the rolling window performed relative to the actual in sample data. Here is the output for the in-sample and out-of-sample estimation.





What in the world is going on here? One could run tests to see if they are “statistically different models” but just a simple eye test could tell you that the out-of-sample is nonrepresentative of the in-sample. This could be due to several factors.

1. The initial model was overfit or not representative. Stock returns in general are said to be a Brownian motion. While the GARCH process can be used to model the volatility a lot of stock returns don’t have this property for the conditional mean. According to the

earlier correlogram in the model selection section, this could be caused by choosing a MA(2) model instead of a random walk. (Highly likely)

2. The second case is unlikely which would be stating that the data set is unrepresentative of the sample. At first daily returns were used and this caused a lot of problems for the estimation models. Using less data points gave us much more clear results and we still have over 1000 data points for the stocks price. (Not likely at all)
3. The third case is that the stochastic algorithm is causing many issues. This is usually caused if one was generating data with a Monte Carlo simulation and this issue would be resolved by running the simulation more times to decrease the variance of the results. This is completely unlikely since there is no data being “simulated” in this analysis. (Not likely at all)

From here the most plausible explanation is the first case which is that the model selection process was poor, and the data can probably be represented by a random walk.

Conclusion:

Overall, the estimation of the VaR seems to have been unsuccessful. When looking back through the analysis one could say that picking a new model would probably be the most ideal method to improving the outcome. One could also try using a Monte Carlo simulation to estimate the VaR however the problem with this method is that there is high chance that the outcome will be much different than the time series approach.