1. *Choose one of the securities to work with. For that security, estimate univariate conditional volatility models of the kind we studied in Topic 2 of the course. Perform statistical analysis to decide which univariate model constitutes the best model for that particular series.*

MSFT returns

Data overview

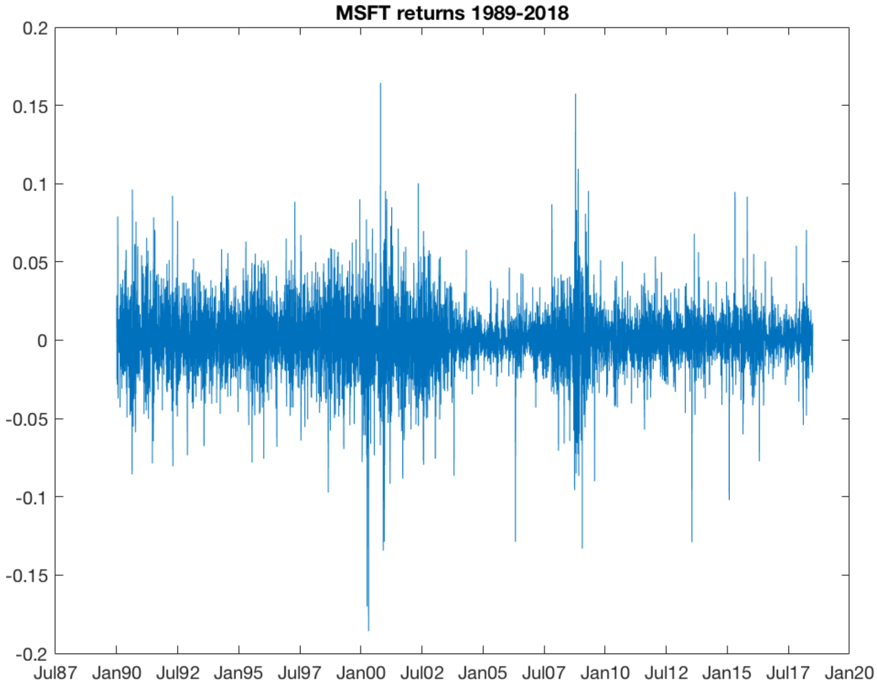

Figure 1 Plot of MSFT daily log-returns

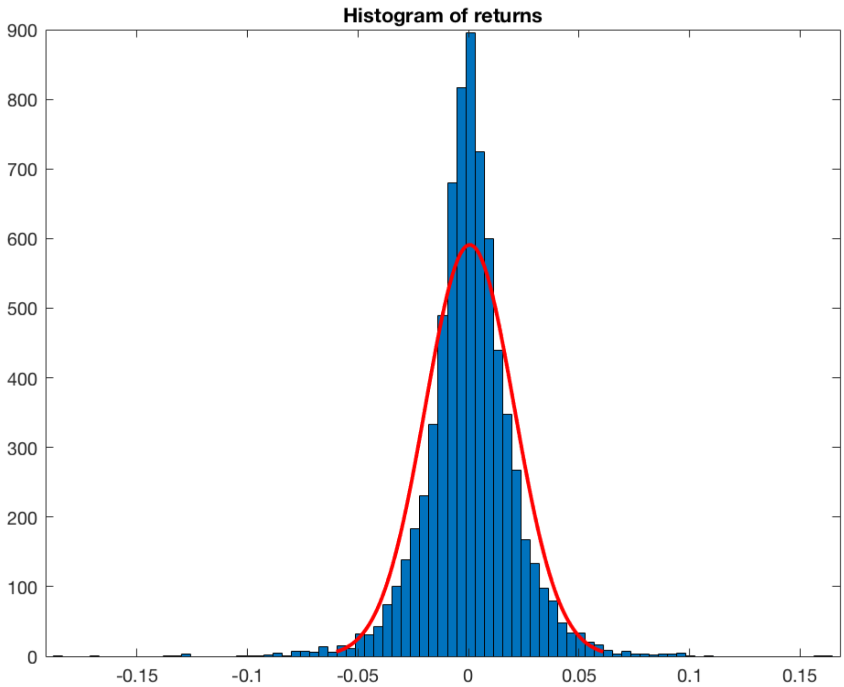

Figure 2 Histogram of MSFT returns

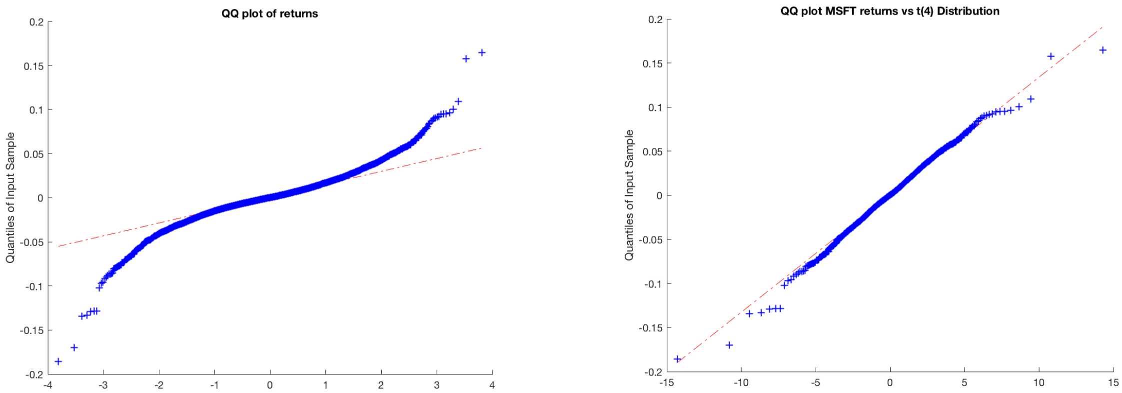

Figure 3 Q-Q plot of MSFT returns vs Standard normal and t(4) Distribution

MSFT Descriptive statistics

Mean	Std. Dev.	Max.	Min.	Skewness	Kurtosis	JB statistic
0.056%	2.01%	16.44%	-18.58%	-0.2216	9.2221	11643
(14.11% Ann.)	(31.91% Ann.)					

- The sample has 7182 observation of MSFT daily log returns from 29 Dec1989 to 29 Jun 2018. In Figure1 daily log-returns are presented. Figure 2 shows a histogram of the data.
- Shown in Figure 3, deviation from linearity and a better fit of t(4) distribution confirms heavy-tailed distribution of MSFT return. With skewness being -0.2216 and Kurtosis being 9.2221, the distribution of MSFT returns is slightly negatively skewed and has fat tails.
- A large Jarque-Bera test statistic suggests that the null hypothesis that the distribution is normal is rejected and there is significant departure from normality of MSFT time series.

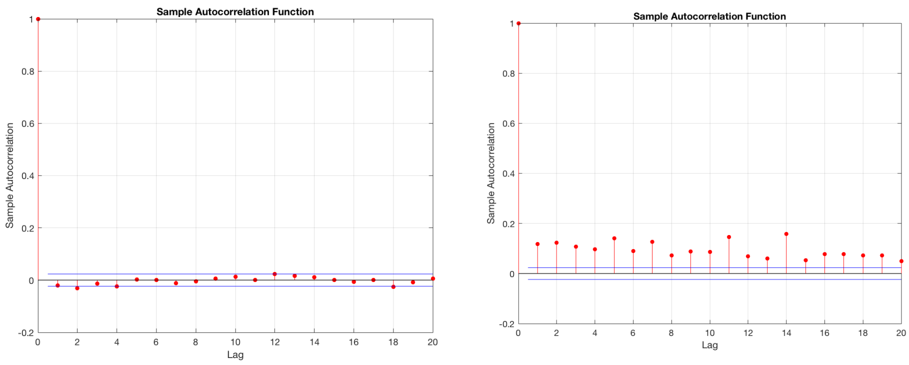

Figure 4 MSFT returns(left) and squared returns(right) autocorrelation

'# Lags'	'Q Stat'	'Crit Val'
1	3.0868	3.8415
2	10.3616	5.9915
3	11.6561	7.8147
4	15.8939	9.4877
5	15.9108	11.0705
6	15.9196	12.5916
7	16.9237	14.0671
8	17.1276	15.5073
9	17.4304	16.9190
10	18.5793	18.3070
11	18.5803	19.6751
12	22.7759	21.0261
13	24.8552	22.3620
14	25.8705	23.6848
15	25.8743	24.9958
16	26.2345	26.2962
17	26.2348	27.5871
18	30.7454	28.8693
19	31.1870	30.1435
20	31.4843	31.4104

Table 1

'# Lags'	'Q Stat'	'Crit Val'
1	101.3057	3.8415
2	209.9380	5.9915
3	293.0438	7.8147
4	360.8494	9.4877
5	504.1459	11.0705
6	563.4187	12.5916
7	679.1474	14.0671
8	717.4518	15.5073
9	774.3766	16.9190
10	828.4388	18.3070
11	981.7868	19.6751
12	1.0152e+03	21.0261
13	1.0410e+03	22.3620
14	1.2199e+03	23.6848
15	1.2403e+03	24.9958
16	1.2836e+03	26.2962
17	1.3279e+03	27.5871
18	1.3650e+03	28.8693
19	1.4035e+03	30.1435
20	1.4216e+03	31.4104

Table 2

Autocorrelation and Predictability

- Figure 4 above shows autocorrelated function of returns and squared returns. Table 1&2 shows test statistics of Ljung-Box test under the null hypothesis that there is no autocorrelation.
- Figure 4 shows weak evidence of autocorrelated returns and strong evidence of autocorrelated squared returns. Table 1 confirms the insignificance of autocorrelated returns at 5% level at lag 1, Q stat= 3.0868<3.84. In Table 2, as all Q statistics are well above critical value, the null hypothesis of no autocorrelation is rejected, presenting strong evidence of serial autocorrelation in squared returns.
- Statistically significant autocorrelation for squared returns suggests volatility clustering and predictability of time series. Different models could be used to model volatility of MSFT returns.

Univariate volatility models

30-day and 60-day Moving Average(MA)

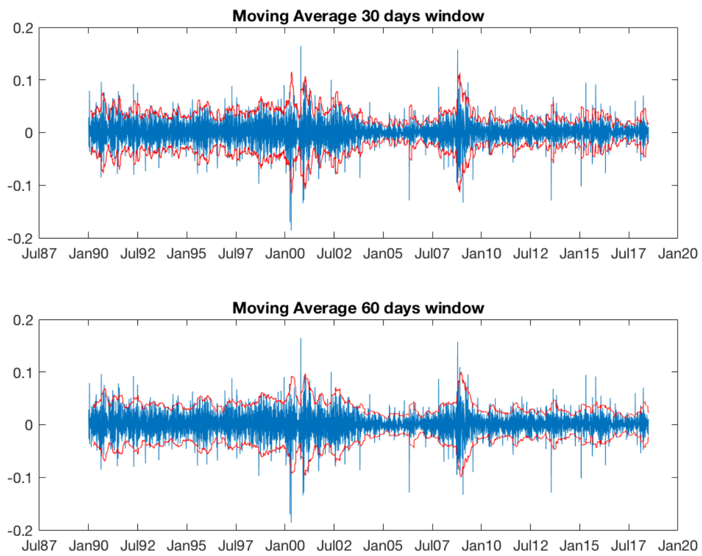

Figure 5 30-day and 60-day moving average

30-day and 60-day Moving Average

- Figure 5 above shows the comparison between 30-day and 60-day Moving Average.
- In a Moving average model current conditional variance estimates are average of equally weighed past squared returns. Different estimation windows can be choose. Shorter time frames prove to be more volatile than longer ones, as plot of estimates is smoother in the 60-days Moving Average in Figure 5. Figure 6 below also demonstrates a wider range of estimates in the 30-day Moving Average at the same period. Such feature makes Moving Average models flexible. However both version of the model shows its poor adaptability of estimates, shown in the figures below, estimates are slowly adjusting to changing volatility.

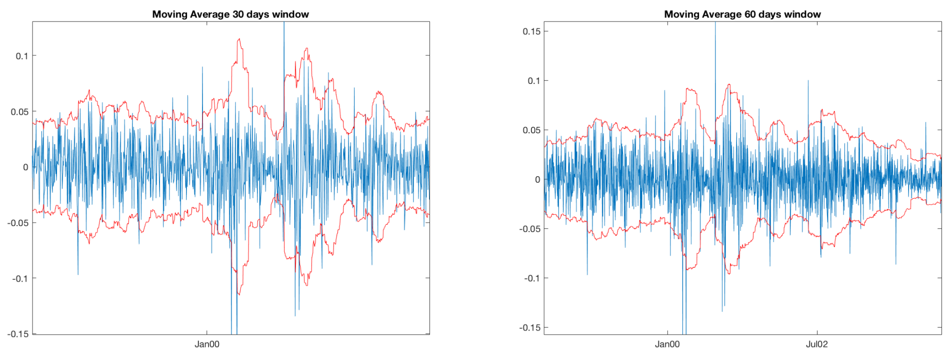

Figure 6 30-day versus 60-day moving average

- Despite desirable features of of MA model, accuracy of volatility forecast is limited and conditional variance estimates are slow at adjusting to changing volatilities. Such specification's equal weighing of past values which may or may not be relevant in the present or into the future is problematic, for it fails to capture the importance of recent influential factors or may include irrelevant information when estimating. Although longer estimation windows spread out more over time periods, general trends may change over time. Therefore, choosing an appropriate window size can be challenging.These inherent limitations undermines the desirability of simple Moving Average models.

Exponentially-Weighted Moving Average(EWMA)

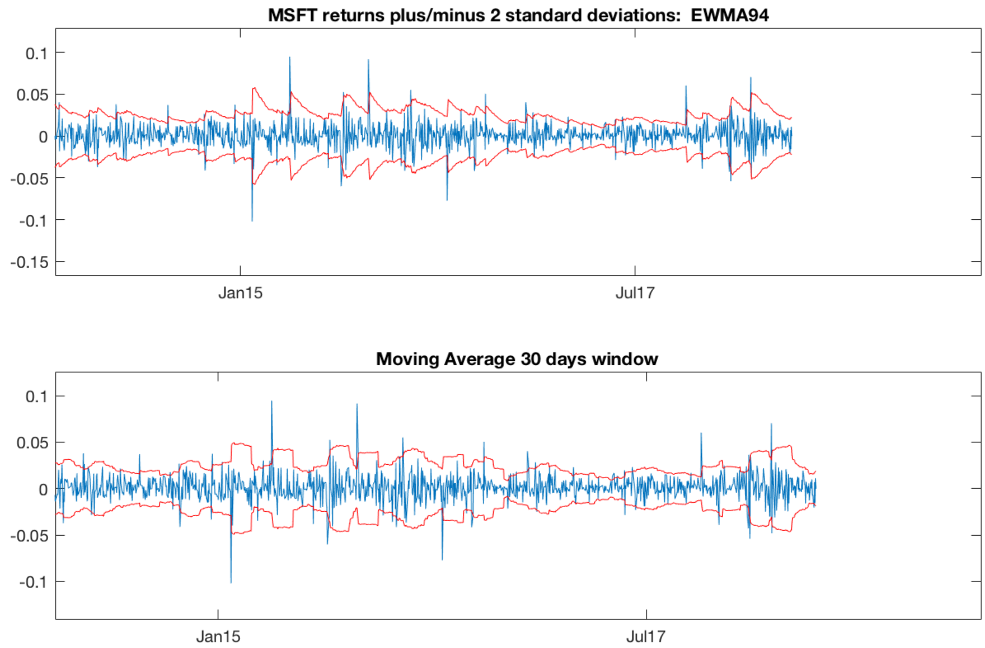

Figure 7 EWMA94 versus 30-day MA

EWMA94 $$\sigma_t^2 = (1-\lambda)r_{t-1}^2 + \lambda\sigma_{t-1}^2$$

- In EWMA, current conditional variance estimates are a weighted average of previous squared returns and conditional volatility estimates, and lamda measures persistence of shocks. Here we assumed 0.94 for daily returns.
- By introducing exponential decay, EWMA model puts more weight on more recent values than distant ones, it improves on the adaptability of estimates. Figure 7 above demonstrates this by comparing volatility estimates produced by EWMA model with a lamda=0.94 and a 30-day MA between 2015 and 2017. It is apparent that between the period of January 2015 and July 2017, 30-day MA volatility estimates remains high after shocks, whereas EWMA94 gradually adapts to the changing volatility, proving improvement on the forecast.
- However, the knowledge of model parameters is insufficient to pin down unconditional variance, as certain transformation of returns also satisfies the defining equation of EWMA process.
- Since we have assumed that mean of returns is zero, using law of iterated expectations, we see, for any future date, conditional expected squared returns equals to conditional expected variance. Thus in EWMA models,

$$\begin{aligned} \mathrm{E}_t[\sigma_{t+k}^2] &= (1-\lambda)\mathrm{E}_t[r_{t+k-1}^2] + \lambda\mathrm{E}_t[\sigma_{t+k-1}^2] \\ &= (1-\lambda)\mathrm{E}_t[\sigma_{t+k-1}^2] + \lambda\mathrm{E}_t[\sigma_{t+k-1}^2] \\ &= \mathrm{E}_t[\sigma_{t+k-1}^2]. \end{aligned}$$

$$\mathrm{E}_t[\sigma_{t+k}^2] = \mathrm{E}_t[\sigma_{t+k-1}^2] = \mathrm{E}_t[\sigma_{t+k-2}^2] = \ldots = \mathrm{E}_t[\sigma_{t+1}^2] = \sigma_t^2.$$

- Consequently, shocks to volatility in EWMA are not mean-reverting, leading to permanently higher variance and a imprecise picture for conditional variance. Such feature owes to the fact that summation of coefficients equals to one.

Indeed, there is improvement in incorporating relevance of past information in EWMA models, as we see adaptability in volatility estimates. EWMA models suffer from drawbacks related to unconditional variance, which could lead to imprecision.

Candidate number 21939

ARCH & GARCH

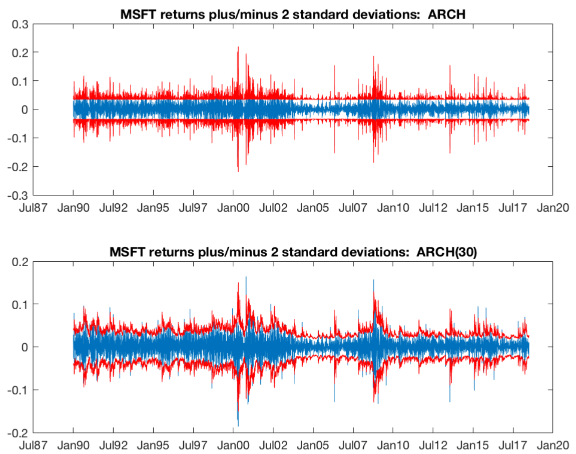

Figure 8 ARCH(1)(upper) & ARCH(30)(lower)

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2$$

	Value	StandardError	TStatistic	PValue
Constant	0.00028643	3.0126e-06	95.078	0
ARCH{1}	0.34123	0.012168	28.044	4.7822e-173

Table 3 Log likelihood 18122.32

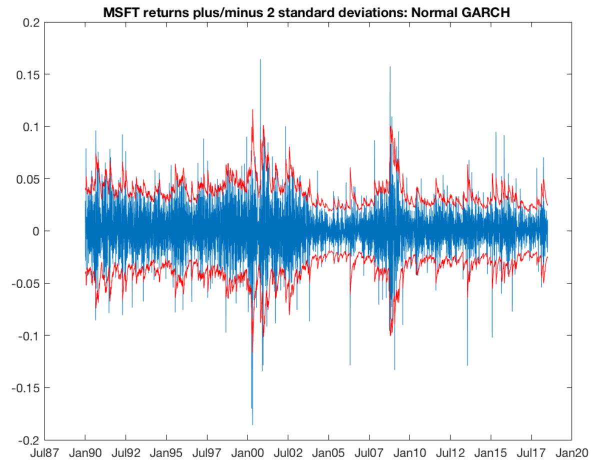

Figure 9 GARCH(1,1)

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2$$

	Value	StandardError	TStatistic	PValue
Constant	3.7293e-06	4.1982e-07	8.8831	6.505e-19
GARCH{1}	0.94178	0.0029306	321.36	0
ARCH{1}	0.04939	0.0022674	21.783	3.361e-105

Table 4 Log likelihood 18655.35

Autoregressive Conditional Heteroskedastic Model (ARCH)

- Table 3 &4 shows estimates of parameters of a ARCH(1) and GARCH(1,1)process, parameter estimates are significant at 1% and 5% level. ARCH models model volatility as a function of past squared returns and past conditional volatility models, they also allow for an arbitrary number N of lags of squared returns.
- Improving on EWMA models, ARCH models effectively incorporate effect of past shock on current volatility estimates and do not suffer from drawbacks related to variance stationarity under certain restrictions(alpha<1). Shocks in ARCH models eventually die out and they replicate fat tails and other features of financial time series even if standardised residuals are normal.
- Upper part of Figure 8 is a plot of volatility estimates using ARCH(1) model. As shown in the figure, the accuracy of volatility estimates are unstable. The model frequently overestimates volatility for the time series, this could be observed at periods from July 2002 and July 2007 and from Jan 2010 to more recent dates. The goodness of fit of ARCH(1) process is therefore limited.

	Value	StandardError	TStatistic	PValue
Constant	6.9717e-05	3.5163e-06	19.827	1.7459e-87
ARCH{1}	0.10625	0.012734	8.3443	7.1632e-17
ARCH{2}	0.10696	0.011543	9.266	1.9322e-20
ARCH{3}	0.053459	0.0073038	7.3194	2.4912e-13
ARCH{4}	0.073753	0.0098076	7.52	5.4776e-14
ARCH{5}	0.042764	0.0093811	4.5585	5.1519e-06
ARCH{6}	0.055382	0.0085956	6.4431	1.1708e-10
ARCH{7}	0.059979	0.010081	5.9497	2.6867e-09
ARCH{8}	0.026238	0.0086588	3.0302	0.0024441
ARCH{9}	0.014672	0.0082092	1.7873	0.073892
ARCH{11}	0.04571	0.010055	4.5462	5.4632e-06
ARCH{12}	0.021751	0.0088645	2.4537	0.014138
ARCH{14}	0.033723	0.0063464	5.3137	1.0744e-07
ARCH{16}	0.031598	0.0084257	3.7503	0.00017665
ARCH{17}	0.0072456	0.0069141	1.0479	0.29466
ARCH{18}	0.0047826	0.0057567	0.83079	0.40609
ARCH{19}	0.021198	0.0061922	3.4233	0.00061855
ARCH{20}	0.0018085	0.0054492	0.33187	0.73998
ARCH{21}	0.0036089	0.0061447	0.58732	0.55699
ARCH{22}	0.018855	0.0077337	2.438	0.014768
ARCH{23}	0.0011513	0.0050259	0.22908	0.81881
ARCH{24}	0.0015997	0.0056933	0.28097	0.77873
ARCH{25}	0.019847	0.0056376	3.5205	0.00043071
ARCH{26}	0.066298	0.0081185	8.1663	3.1802e-16
ARCH{27}	0.013385	0.0067576	1.9807	0.047627
ARCH{28}	0.0064731	0.0079149	0.81784	0.41345
ARCH{29}	0.021221	0.0061563	3.447	0.00056675
ARCH{30}	0.0081709	0.0072062	1.1339	0.25685

Table 5 Log likelihood 18686.1537724795

•Shown in Table 5 on the left, coefficients are significant at 0.1% up to 7 lags, statistical insignificance starts to appear after that. Unconditional variance for the process exits, as coefficients sum to smaller than 1. It is also worth noting that including more lags often leads to estimation difficulties and overfitting. However, such insignificance does not necessarily undermines improvements upon ARCH(1).

•Problem of inaccuracy of estimates is alleviated when more lags are included in the model. We can observe the improvement in the lower part of Figure 8. For the same time periods mentioned above, estimates produced by ARCH(30) fits the time series better.

•LR test statistics for ARCH(1) and ARCH(30) follows a distribution of chi square with 29 degrees of freedom (29 restrictions). With LR=2(18686.15 -18122.32)=1127.66 >42.5570, the null hypothesis of constraints hold is rejected at 5% level.

ARCH(1) versus GARCH(1,1)

- GARCH models are generalised version of of ARCH. The model includes an additional term of past variance. Alpha corresponds innovations(effect of news on conditional variance), which is particularly important when modelling volatility of a technology company in an environment that is subject to technological changes that have huge impact to the financial market. Beta corresponds to memory which lead to persistence in volatility estimates.
- Table 4 presents parameter estimates of GARCH(1,1) for the time series. Alpha+Beta= 0.94178+0.04939 = 0.99117<1, thus unconditional variance of the process is finite.
- GARCH(1,1) carries same benefits of ARCH models by construction, and generally leads to more stable volatility estimates for the time series. Shown in Figure 10 below, a normal GARCH(1,1) proves to produce less volatile volatility estimates than both ARCH.
- LR test statistic follow a distribution of chi squared with 1 degree of freedom. With LR=2(18655.35-18122.32)=1066.06>3.84, the null hypothesis that the constraint holds is rejected at 5%.

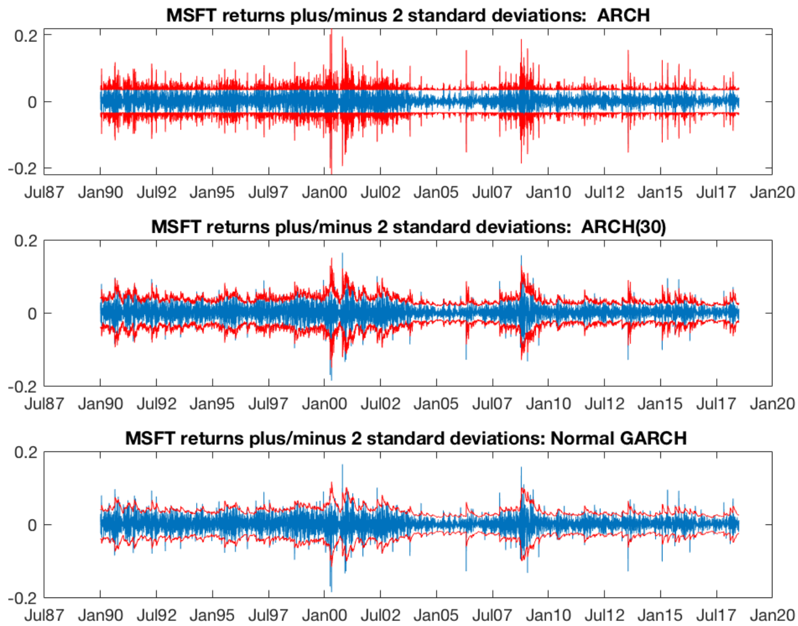

Figure 10 ARCH(1),ARCH(30)&GARCH(1,1)

AIC(Akaike Information Criterion) and BIC (Bayesian Information Criterion)

- Both AIC and BIC provide a mean for model selection by estimating relative loss of information by a given model. Two criterions are similar, but use different penalty for number of parameters, 2k in AIC and ln(n) in BIC, with k being number of parameters and n being sample size. Table below shows results of AIC and BIC for the three models.
- For model comparison, the lowest score is preferred. AIC suggests ARCH(30) and BIC suggests GARCH(1,1).
- Statistical inference suggests that ARCH(30) and GARCH(1,1) has relatively similar loss of information. But with more parameters to estimate, ARCH(30) also face the issue of overfitting. By introducing a greater penalty in BIC, GARCH(1,1) is preferred over ARCH(30), suggesting overfitting of ARCH(30).

Models	AIC	BIC
ARCH(1)	-36240.63	-36226.87
ARCH(30)	**-37316.31**	-37123.69
GARCH(1,1)	-37304.70	**-37284.06**

Table 6

Residual analysis

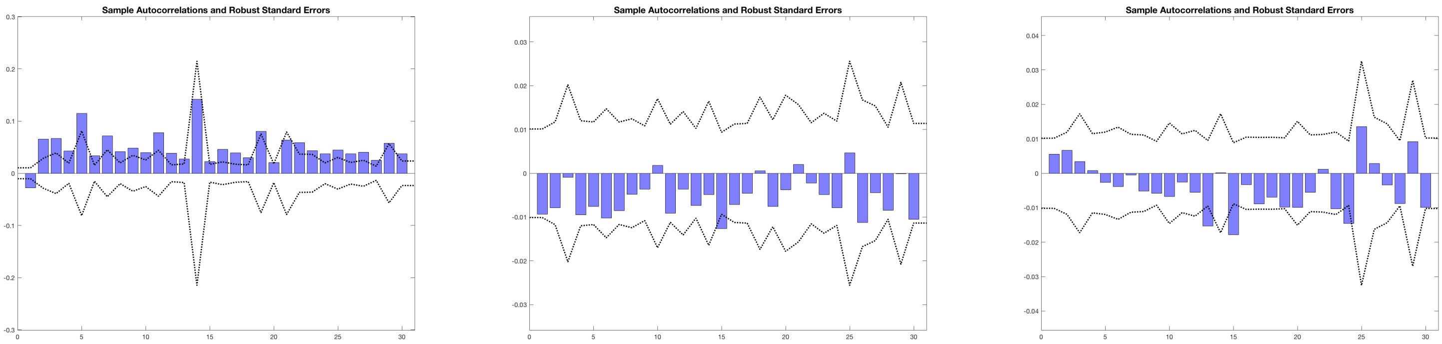

Figure 11 ACF of squared residuals of ARCH(1)(left), ARCH(30)(middle) &GARCH(1,1)(right)

Lags'	'Q Stat'	'Crit Val'
	5.6294	3.8415
	36.5921	5.9915
	68.5040	7.8147
	81.7354	9.4877
	176.4369	11.0705
	184.6415	12.5916
	221.7803	14.0671
	234.0728	15.5073
	251.0234	16.9190
	262.2067	18.3070
	305.7988	19.6751
	316.3465	21.0261
	321.7799	22.3620
	465.9771	23.6848
	469.7035	24.9958
	484.6394	26.2962
	495.7795	27.5871
	502.2499	28.8693
	548.6887	30.1435
	551.7914	31.4104

'# Lags'	'Q Stat'	'Crit Val'
1	0.6252	3.8415
2	1.0727	5.9915
3	1.0784	7.8147
4	1.7234	9.4877
5	2.1384	11.0705
6	2.8855	12.5916
7	3.4067	14.0671
8	3.5736	15.5073
9	3.6696	16.9190
10	3.6922	18.3070
11	4.2887	19.6751
12	4.3841	21.0261
13	4.7725	22.3620
14	4.9459	23.6848
15	6.1032	24.9958
16	6.4693	26.2962
17	6.6192	27.5871
18	6.6211	28.8693
19	7.0364	30.1435
20	7.1389	31.4104

'# Lags'	'Q Stat'	'Crit Val'
1	0.2111	3.8415
2	0.5213	5.9915
3	0.5992	7.8147
4	0.6038	9.4877
5	0.6583	11.0705
6	0.7675	12.5916
7	0.7695	14.0671
8	0.9665	15.5073
9	1.2143	16.9190
10	1.5428	18.3070
11	1.5919	19.6751
12	1.8122	21.0261
13	3.5055	22.3620
14	3.5058	23.6848
15	5.8004	24.9958
16	5.8785	26.2962
17	6.4516	27.5871
18	6.7987	28.8693
19	7.4862	30.1435
20	8.1919	31.4104

Table 7 LB test statistics ARCH(1),ARCH(30)&GARCH(1,1)(left to right)

- Figure 11 above shows strong evidence of autocorrelated squared residuals from ARCH(1) process, which is inconsistent with our assumption that residuals are i.i,d, and very weak evadence of autocorrelation in that from ARCH(30) and GARCH(1,1) process.
- Table 7 shows LB test statistics for testing existence of autocorrelation in squared standardised residuals produced by three models. The table confirms autocorrelation of standardised squared residuals from ARCH(1) and and no autocorrelation in squared residuals from ARCH(30) and GARCH(1,1).

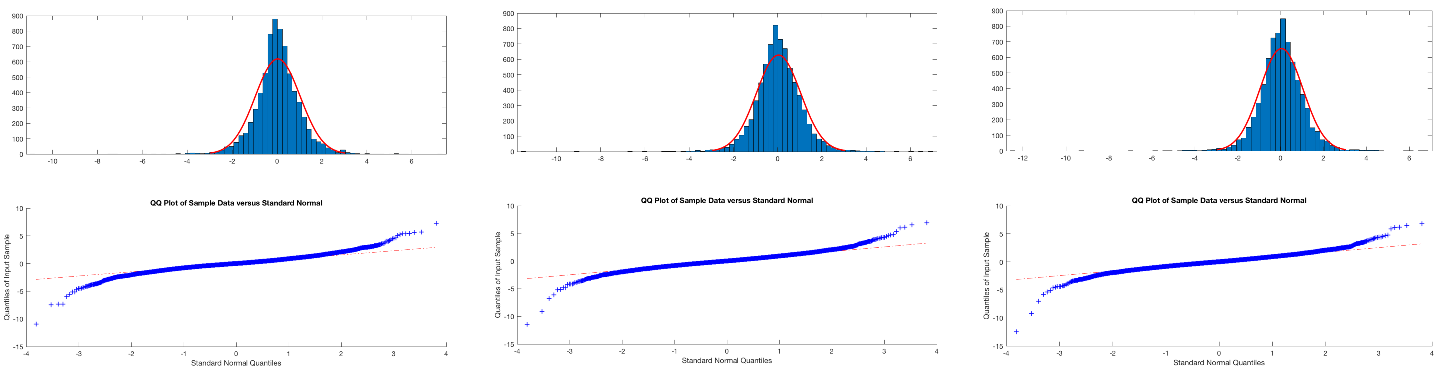

Figure 12 plot of standardised residuals of ARCH(1)(left), ARCH(30)(middle) &GARCH(1,1)(right)

Models	**Skewness**	**Kurtosis**	**JB test statistic**
ARCH(1)	-0.2757	9.1637	11460.04
ARCH(30)	-0.2034	9.1393	11328.65
GARCH(1,1)	-0.2938	10.2580	**15867.56**

- Figure 12 above plots the standardised residuals from the three model. All three series of residuals have heavy tails, as Q-Q plots show deviation from linearity. JB test statistics conclude that the distribution of all three series of standardised residuals are significantly different from being conditionally normal, GARCH(1,1) departure from normality is the most significant.
- It safe to conclude that both ARCH(30) and GARCH(1,1) are preferred to ARCH(1), given ARCH(1)'s autocorrelated squared residuals, limited accuracy and greater loss of information. Both having no serial autocorrelation in squared residuals and similar loss of information, performance of ARCH(30) and GARCH(1,1) is comparable.

GARCH extended

GJR-GARCH

	Value	StandardError	TStatistic	PValue
Constant	7.9599e-06	3.4463e-07	23.097	4.9688e-118
GARCH{1}	0.89166	0.005006	178.12	0
ARCH{1}	0.065513	0.005427	12.072	1.4882e-33
Leverage{1}	0.065596	0.0076856	8.535	1.4019e-17

Table 9 Log likelihood 18655.35

GARCH(2,1)

	Value	StandardError	TStatistic	PValue
Constant	4.5894e-06	7.8277e-07	5.8631	4.5428e-09
GARCH{1}	0.63094	0.1509	4.1811	2.9016e-05
GARCH{2}	0.29627	0.1436	2.0632	0.039094
ARCH{1}	0.061782	0.0065089	9.4919	2.2683e-21

Table 10 Log likelihood 18657.54

GARCH t-distribution

	Value	StandardError	TStatistic	PValue
Constant	1.4062e-06	5.7163e-07	2.4599	0.013897
GARCH{1}	0.93969	0.0049097	191.39	0
ARCH{1}	0.060311	0.0053416	11.291	1.4551e-29
DoF	5.1248	0.27311	18.764	1.4796e-78

Table 11 Log likelihood 19058.67

GARCH(2,1) t-distribution

	Value	StandardError	TStatistic	PValue
Constant	1.6788e-06	7.4428e-07	2.2556	0.024094
GARCH{1}	0.63748	0.23793	2.6793	0.0073776
GARCH{2}	0.28759	0.22548	1.2754	0.20215
ARCH{1}	0.074931	0.013962	5.3667	8.0171e-08
DoF	5.1405	0.27351	18.794	8.3912e-79

Table 12 Log likelihood 19060.40

GJR-GARCH

- A limitation of GARCH(1,1) is that it does not incorporate leverage effect which is among the stylised facts of financial returns. Leverage effect considers that negative returns at time t-1 have stronger impact on variance at time t than positive returns.
- Table 9 on the left shows all parameters are statistically significant, including the parameter estimate for leverage effect.
- Log likelihood ratio test is conducted to test significance of leverage effect. LR test for GJR-GARCH and GARCH(1,1) follows a chi square distribution of 1 degree of freedom. With LR = 2(18655.35-18655.35)= 0.00<3.84, the null hypothesis that the constraint holds is not rejected. We conclude that the leverage effect is statistically insignificant.

GARCH(2,1)

- Extending on GARCH(1,1), GARCH(2,1) process adds an additional lag of sigma squared.
- Table 10 on the left shows all parameters are statistically significant. However, parameter for GARCH{2} is only significant at 5% level. With LR=2(18657.54-18655.35)=4.38>3.84, the null hypothesis that the constraint holds is rejected at 5% level. Parameter estimate for the second lag of past conditional variance is statistically significant at 5% level.

GARCH Student-t distribution

- Having observed heavy tails of standardised residuals of GARCH(1,1), instead of conditionally normal, we assume that standardised residuals follows a student t distribution.
- Parameter estimates are presented in Table 11, all parameter estimates are significant at 5% level. However, the process has infinite unconditional variance, as sum of coefficients greater than 1. Increasing order of GARCH with t-distribution does not resolve the instability and parameter estimate for GARCH{2} is insignificant. Such specification does not fit the time series well.

Residual analysis

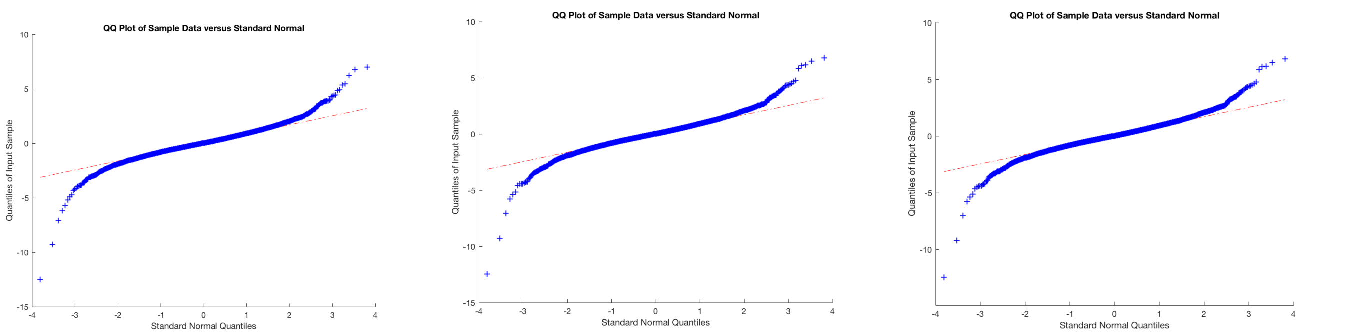

Figure 13 QQplot of standardised residuals of GJR-GARCH(1)(left), GARCH(2,1)(middle) &GARCH(1,1)(right)

Models	Skewness	Kurtosis	JB test statistic
GJR-GARCH	-0.2984	10.6136	17453.27
GARCH(2,1)	-0.2998	10.2860	15993.58
GARCH(1,1)	-0.2938	10.2580	15867.56

- In Figure 13, it is also demonstrated by the QQ-plots that the residuals exhibit heavier tails for GJR-GARCH and GARCH(2,1).
- Descriptive statistics show that residuals for GARCH models are negatively skewed and have fat tails. Large JB test statistic shows departure from normality. GJR-GARCH departs even further from normality than normal GARCH.

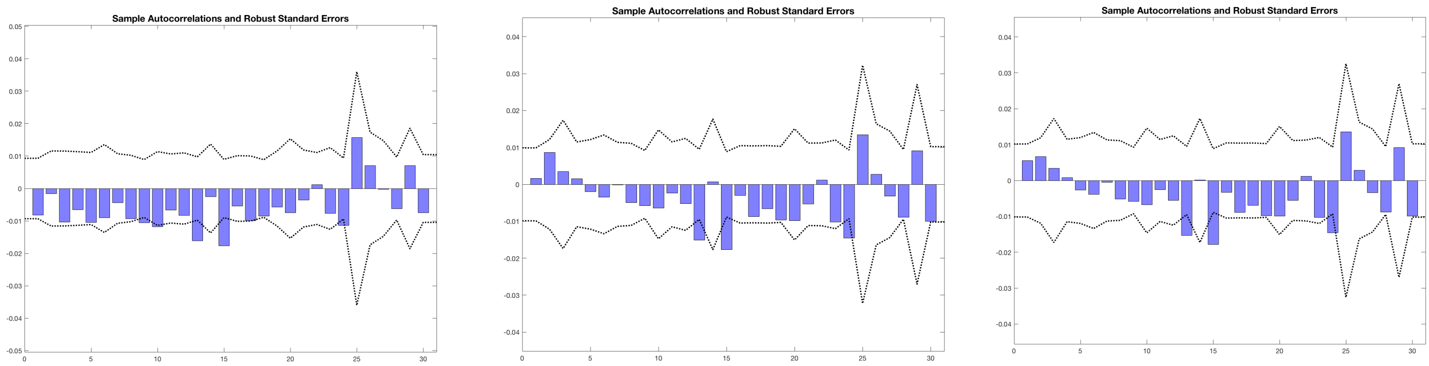

Figure 14 ACF of squared standardised residuals of GJR-GARCH(1)(left), GARCH(2,1)(mid),GARCH(1,1)(right)

'# Lags'	'Q Stat'	'Crit Val'
1	0.4878	3.8415
2	0.5050	5.9915
3	1.2759	7.8147
4	1.5815	9.4877
5	2.3670	11.0705
6	2.9500	12.5916
7	3.0877	14.0671
8	3.7105	15.5073
9	4.5200	16.9190
10	5.5280	18.3070
11	5.8439	19.6751
12	6.3446	21.0261
13	8.2205	22.3620
14	8.2675	23.6848
15	10.5068	24.9958
16	10.7185	26.2962
17	11.4247	27.5871
18	11.9445	28.8693
19	12.1812	30.1435
20	12.5825	31.4104

'# Lags'	'Q Stat'	'Crit Val'
1	0.0198	3.8415
2	0.5604	5.9915
3	0.6489	7.8147
4	0.6660	9.4877
5	0.6935	11.0705
6	0.7784	12.5916
7	0.7786	14.0671
8	0.9514	15.5073
9	1.1908	16.9190
10	1.4876	18.3070
11	1.5273	19.6751
12	1.7218	21.0261
13	3.3653	22.3620
14	3.3689	23.6848
15	5.6230	24.9958
16	5.6889	26.2962
17	6.2383	27.5871
18	6.5536	28.8693
19	7.2223	30.1435
20	7.9210	31.4104

Table 13 LB test statistics GJR-GARCH(left)&GARCH(2,1)(right)

- Figure 14 above shows strong evidence of no autocorrelation is squared standardised residuals from the two models.
- Table 13 shows LB test statistics confirms no autocorrelation in squared standardised residuals for the two models, as all Q stats are below critical value.
- AIC BIC scores of four models ARCH(30), GJR-GARCH,GARCH(1,1) & GARCH(2,1) are presented in Table 14 below. ARCH(30) slightly outperforms three GARCH models in AIC. But with greater penalty for including more parameters, GARCH(1,1) is preferred to ARCH(30), as it has the lowest BIC score.

Table 14

Models	AIC	BIC
ARCH(30)	**-37316.31**	-37123.69
GJR-GARCH	-37302.71	-37275.19
GARCH(1,1)	-37304.70	**-37284.06**
GARCH(2,1)	-37307.09	-37279.57

Summary

Models	Constant	ARCH[1]	GARCH[1]	GARCH[2]	Leverage	DoF
GJR-GARCH	0.0000079599***	0.065513***	0.89166***		0.065596***	
GARCH(1,1)	0.0000037293***	0.04939***	0.94178***			
GARCH(2,1)	0.0000045894***	0.061782***	0.63094***	0.29627*		
GARCH-t	0.0000014062*	0.060311***	0.93969***			5.1248***
GARCH(2,1)-t	0.0000016788*	0.074931***	0.63748**	0.2876		5.1405***

***= Statistically significant on 0.1% level
**= Statistically significant on 1% level
*= Statistically significant on 5% level

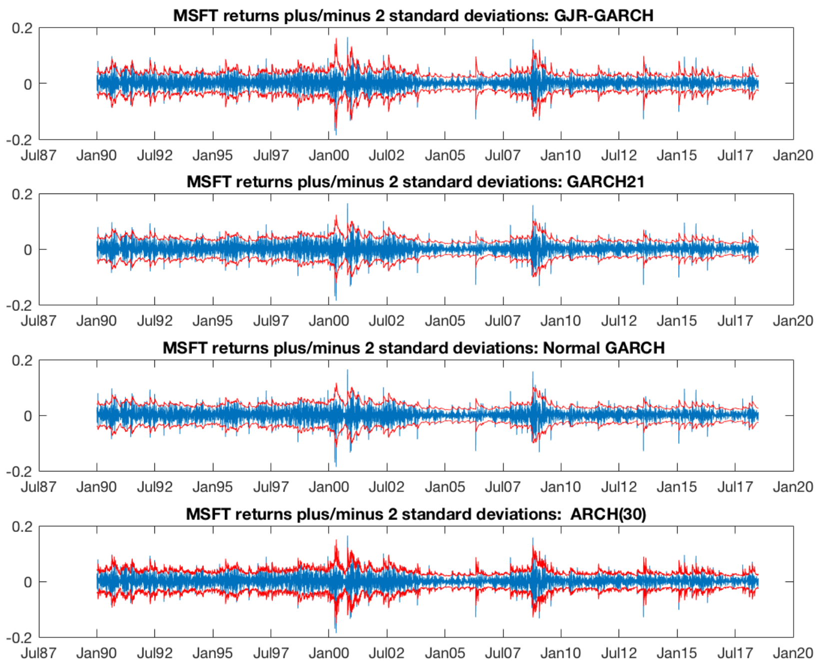

Figure 15 Plot of all 4 models

Conclusion

Among all models discussed, MA and EWMA suffers major drawbacks such as lack of adaptability and non-stationarity, which both could result in inaccurate estimates for volatility, seriously undermining the desirability of the model. Simplicity of such models is helpful for other purposes.

Improving on MA models, (G)ARCH models produces much more reasonable estimates. Both ARCH(30) and GARCH(1,1) are significantly preferred to ARCH(1). Extending on GARCH(1,1), we observe that leverage effect is statistically insignificant and GJR-GARCH does not have the lowest score in both AIC and BIC, and unconditional variance does not exist for student-t GARCH. When comparing ARCH(30) and the GARCH family, performance of ARCH(30) is not dominant, despite its lowest AIC score. Although large sample size alleviates estimation difficulty, ARCH(30) is still penalised for its overfitting and produces more volatile estimates relative to GARCH, shown in Figure 15. Given analysis above, including more parameters in models does not significantly improve models' performance when fitting MSFT series. We therefore conclude that a GARCH(1,1) would best fit the time series.