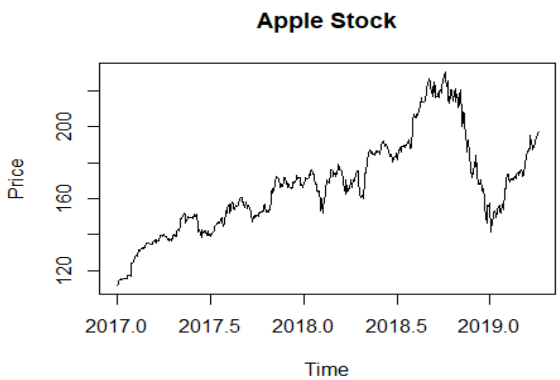
Applied Time Series Analysis: Final Project Report

Jonathan Mac

Introduction

Background: The stocks dataset contains stock prices for 6 Tech giants, one of which I will analyze in this report. Apple is public company renowned for its iphone innovations, and has grown tremendously in the past decade. Its stock is an attractive equity investment with large trading volumes. Building a model to predict market movement based on historic data improves investing decisions in real time.



Data at first look: Apple's stock data represents daily closing prices over a roughly 2 year period from 2017 to 2019. Markets operate only during business, so one typical week equals 5 trading days, and one year equals 251 trading days. All data is full and accurate, no missing values here.
I observe a steady, linear upward trend in Apple's stock price from 2017 to 2018.5. From 2018.5 to 2018.75, the price spikes but then drops abruptly until 2019 when it recovers. This reversal point in 2019 was either due to gravity toward fundamental market price or an external force helping the company recover after a crisis.

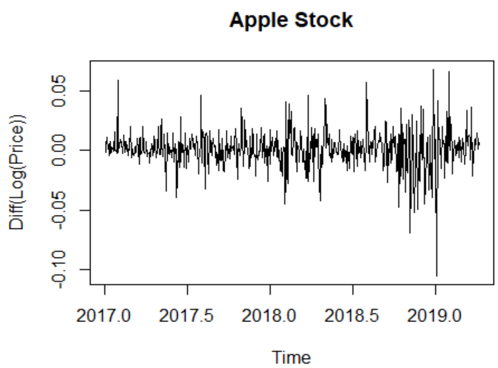
Potential Models: In brief, this time series has choppy trends and erratic behavior, so its mean and standard deviation is non-zero, implying non-stationarity. The trends do not resemble a seasonal pattern either, as stock returns follow random movement. The non-stationarity and volatility clustering can be remedied by transformations (log, differencing).

Since stock returns are correlated, fitting an Autoregressive model helps capture the effects of past prices. I will include drift because of the apparent trends. Additionally, I will fit a Generalized Autoregressive Conditionally Heteroscedastic (GARCH) Model to account for the volatility clustering and non-stationarity of the data. A state space model will also help reveal how Apple stock is affected by other companies in the data.

Insights to look out for:
Fitting a robust model can help:
* forecast stock prices/returns into the near future
* assess volatility during certain times, which helps risk-averse investors decide whether to sell their stake in Apple for safer investments.
* identify trends in stock prices/returns
* analyze how other companies' stock movement affects Apple's (only applicable to state space model)

Autoregressive Integrated Moving Average (ARIMA) Model

Transforming Data: Taking the natural log of the stock prices helps smoothe our bumpy data. As a next step, I took the first difference of the log(price) to account for the linear upward/downward trends. No further differencing is needed because the trends in the graph, though choppy, have been mainly linear. Based on theory, the diff(log(price)) approximately represents the stock return. After transforming, the data doesn't look perfectly like white noise, but close.



Overview: The ARIMA(p,q) model can take many combinations of p and q, even solely pure AR(p) models (autoregressive; when q = 0) or pure MA models (Moving Average; when p = 0). Ultimately, the best model fit will be an ARIMA(2,1,2) model.

The model specification is $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2}$

Determining the order of ARIMA(p,q) model: The Autocorrelation and Partial Autocorrelation functions help determine the order of our models. As seen in figure 1 (refer to appendix), the 7th lag goes above our both ACF/PACF cut-offs and may be significant. Potential models include all combinations of p = 1…7 and q = 1…7 for ARMA(p,q), AR(p), and MA(q).

<u>*Checking Information Criterion*</u>

One way to determine order (p,q) that yields the best fitting ARIMA model is through an Information Criterion. After implementing a for-loop and iterating through:

- p = 1…7 for AR(p)
- q = 1…7 for MA(q)
- every combination of (p,q) from p = 1…7 and q = 1…7 for ARIMA(p,q)

I obtained information criterion values which included AIC, AICC, and BIC for each model type and its order combinations (see figure 2 in appendix for more detail). For each model type, I determined all 3 minimum information criterion (top) and their respective model orders (bottom).

Information Criterion Values	*Lowest AIC*	*Lowest AICC*	*Lowest BIC*
AR(p = 1…7)	-3094.903	-3094.882	-3086.223
MA(q = 1..7)	-3094.91	-3094.889	**-3086.23**
ARMA(p = 1…7, q = 1…7)	**-3097.772**	**-3097.665**	-3080.174

Corresponding Model Orders	*Lowest AIC*	*Lowest AICC*	*Lowest BIC*
AR(p)	(p = 1)	(p = 1)	(p = 1)
MA(q)	(q = 1)	(q = 1)	**(q = 1)**
ARMA(p = 1…7, q = 1…7)	**(p = 2, q = 2)**	**(p = 2, q = 2)**	(p = 1, q = 1)

The ARMA(2,2) model has the lowest AIC and AICC values, while the MA(1) model has the lowest BIC value. The reason is that BIC favors sparser models by more heavily penalizing models with a greater number of parameters. The ARMA(2,2) model wins by majority vote since two of its three information criterion values are lowest.

Parameters of ARIMA(2,1,2) Model:

```
## Coefficients: (see figure 3 in appendix for more detail)
##          ar1      ar2      ma1     ma2    mean
##        0.920  -0.9895  -0.9375  0.9861  5e-04
## s.e.  0.009   0.0119   0.0124  0.0166  7e-04
```

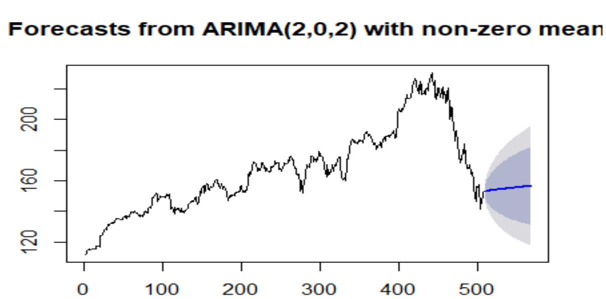
mean = 5e-04 represents the drift. So the mean annual return (1 year has 251 trading days) is 5e-04*251 = 0.1255 or 12.55%. This is a credible because indexes like the dow jones averages a 10% annual return. Apple is an elite growth company with a history of superior returns based on the past decade of data. So a 12.55% annual return estimate from two years of data is reasonable.

AR coefficients ar1 = 0.920 and ar2 = -0.9895 imply that yesterday's stock price positively influences today's, but stock price two days ago negatively influences today's price. Optimism is contagious, so it makes sense that yesterday's bull run promotes a positive outlook on the stock today. And the vice versa holds for a bearish scenario.

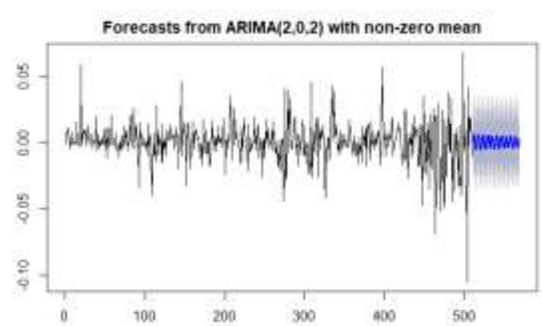
But the negative ar2 term is harder to speculate, just like the positive-negative pair for ma1 = -0.9375 and ma2 = 0.9861. Nonetheless, this confirms the notion that markets are truly random, and models can't always accurately explain trading behvaior or predict movements.

Forecast:

<u>Price</u>



<u>Returns</u>



Investors react to earnings announcements, which occur quarterly or every 3 months. So it's sensible to forecast prices for each quarter, which has 60 days. I remove one day for the next earnings announcement.

The forecasts doesn't answer all our initial questions. For instance, price forecasts (left) predicts an upward trend, but not volatility. The blue area represents the 95% confidence level forecast, which is round and expands infinitely. Stock prices will be difficult to predict with a pure ARMA model because of such a wide interval. So the lengthier the forecast, the poorer the forecast because the ARIMA model doesn't forecast volatility. Additionally, forecasted returns (right) narrowly oscillates slightly above 0%, but the blue area is sizably larger indicating huge volatility. With just an ARIMA model, analysts will never know during which periods will be most risky.

I removed the last 59 observations to test the prediction accuracy of the ARIMA(2,1,2) model. The mean squared error (MSE) calculated with the 59 forecasted values and the last 59 observations was 0.0002525792, or .02525%. This means the ARIMA model forecasts daily returns that are off on average by .02525%. Though the expected error (MSE) is very minute, the volatility is wild, so this ARIMA(2,1,2) model has a lot of short-comings.

Generalized Autoregressive Conditionally Heteroscedastic (GARCH) Model

Overview: The GARCH model will help model non-stationary nature of Apple's stock time series by capturing and accounting for the periods of fluctuating volatility. I will fit a GARCH(1,1) model with specification $\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2 + \beta_1 \sigma_{t-1}^2$.

Fitting an AR(1) model, $x_t - \mu = \phi_1(x_{t-1} - \mu) + w_t$, is also helpful because stock prices are correlated and exhibit trend behavior. Previously, I differenced the log of the stock data to account for several linear trends, so I will include a drift component in the AR(1) model. My final model combines AR(1) + GARCH(1,1).

Model Fit: I check the qqPlot to assess the model fit of **AR(1) + GARCH(1,1)**. The tails stray away from our normality line, indicating heavier tails (see figure 4 in appendix). Instead of a normal distribution, Student's t distribution may yield a better fit.

I plot the qqplot of my AR(1) + GARCH(1,1) model under the student's t distribution, with an estimated 4 degrees of freedom (see figure 5 in appendix). The sample quantiles near the tails fit much better, so I will AR(1) + GARCH(1,1) with Student's t distribution.

Checking Information Criterion
AR(1) + GARCH(1,1)
Akaike: -5.6238 Bayes: -5.5822

AR(1) + GARCH(1,1) with student's t distribution
Akaike: -5.8015 Bayes: -5.7515

The AIC and BIC values are lower in the Student t distribution case, indicating that Student's t distribution is a more appropriate assumption than that of a normal distribution.

What about other orders AR(p) and GARCH(p,q)?
Another way to determine the order of the GARCH(p,q) is to plot the ACF/PACF of the squared transformed data, and determine the significant lags. As shown in figure 6 (see appendix) lag = 5 is significant for both ACF/PACF. But after checking information criterion values for GARCH(p = 2…5, q = 2…5), GARCH(1,1) is still a superior fit since its AIC and BIC values are still lowest.

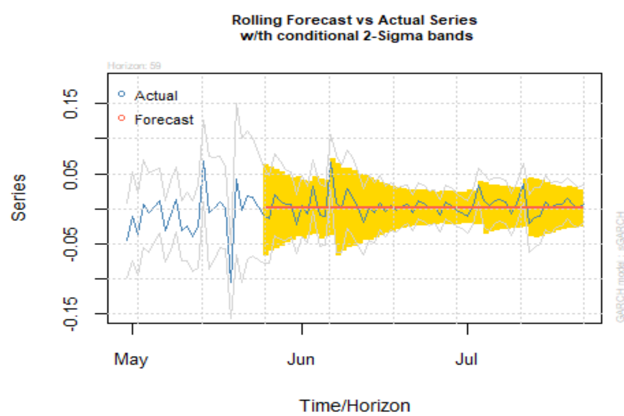
Parameters of AR(1) + GARCH(1,1) Model:
Let's interpret the parameters from our model specification $\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2 + \beta_1 \sigma_{t-1}^2$.

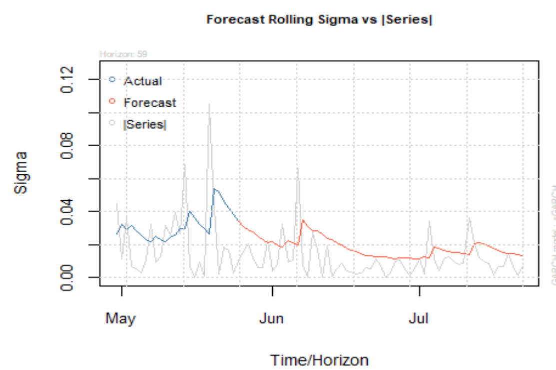
```
Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.001254    0.000471  2.66112 0.007788
ar1     0.008114    0.045276  0.17922 0.857767
omega   0.000017    0.000007  2.38113 0.017260
alpha1  0.209014    0.067696  3.08754 0.002018
beta1   0.789986    0.067283 11.74125 0.000000
shape   3.120591    0.519188  6.01052 0.000000
```

mu = 0.001254 represents the daily mean return. So the annual mean would be 0.001254*251 = 0.314754, or 31.47%. This makes sense because 10 years ago, Apple stock was $30. Currently, it's almost $300. So this ~30% return is a reasonable estimate just based on two years of data. The p-value of mu is .007 which is less than our 5% confidence level, indicating that this mu is significant.

beta1 = 0.789986 is the coefficient of the variance lag terms such as σ_{t-1}^2. Since beta1 is near 1, the volatility of the first lag heavily influences that of the next. The p-value of nearly 0 also indicates signifiance that there are trailing periods of volatilties, so stock volatility is influenced by fluctuations in the past.

Forecasting: I perform a rolling forecast, which predicts 1-step-ahead incrementally while incorporating new information until n-steps are done. Consistent with my previous logic, I decide to forecast 59 trading days into the future because investor sentiment changes for every quarterly earnings announcement which is 3 months or 60 trading days. Below are forecasts of the data and σ_t





The 2-Sigma bands imply a 95% confidence level. The confidence bands shaded in yellow capture the actual last 59 days almost perfectly, except for some miniscule parts of the upward spikes. Thanks to GARCH model they now even forecast clusters of volatility where bursts occur in late May, early June, and Mid July. Earnings announcement happens in June but volatility is relatively stable, indicating that Apple's performance probably met investor's expectations. Otherwise, volatility would have increased then because of surpassed or lackluster expectations.

The rolling forecast of the σ_t shows that our AR(1) + GARCH(1,1) was less volatile than the σ_t of the series. This indicates that our model is more conservative about the volatility of the stocks. A meaningful interpretation of this is that Wall Street Analysts can pitch Apple as a stable stock investment because this model still forecasts prices well. On the other hand, an overestimated σ_t would harm Apple because its stock would be considered riskier than expected. It would make sense for analysts to stick with the more conservative model for recommendations.

I used a built-function fpm which helped solve for the mean square error of 2.183728e-04, or .021837%. Since I performed a rolling forecast of 10 steps, I referred to the 11th column of the fpm function output for the MSE that corresponded to my 10th rolling forecast. The MSE means that the error of our daily forecast returns is .021837%.

Lastly, the proportion of observations that lied outside the 95% confidence yellow region was 0. The GARCH model does a good job forecasting for the next 59 days with rolling forecast.

State Space Model

Overview: I fit a multivariate state space model to all 6 stock price series including Apple, Adobe, Amazon, Facebook, Google, and ^GSPC, but mainly focusing on Apple. I removed the last 59 observations to test the forecast accuracy later. Apple does not have any seasonal patterns, and its movements are choppy so there is no one-sided trend.

State Equation: $x_t = \phi x_{t-1} + w_t$
Observation Equation: $y_t = A_t x_t + v_t$

where x_t is a 6x1 matrix representing all 6 companies, and ϕ is a 6x6 matrix representing the effects of lags and the relationship among the companies. y_t is just a noiser version of x_t, both with standard normal error terms w_t and v_t.

The residual plots of state and observation equation in figures 7.3 and 7.4 respectively show clusters and patterns, not representative of white noise. The qqplots in figures 7.5 and 7.6 exhibit heavy tail-end behavior, implying that none of the series are normal.

Parameters State Equation of Apple:

State Equation of Apple:
$x_{t,1} = \phi_{1,1}x_{t-1,1} + \phi_{1,2}x_{t-1,2} + \phi_{1,3}x_{t-1,3} + \phi_{1,4}x_{t-1,4} + \phi_{1,5}x_{t-1,5} + \phi_{1,6}x_{t-1,6} + w_t$

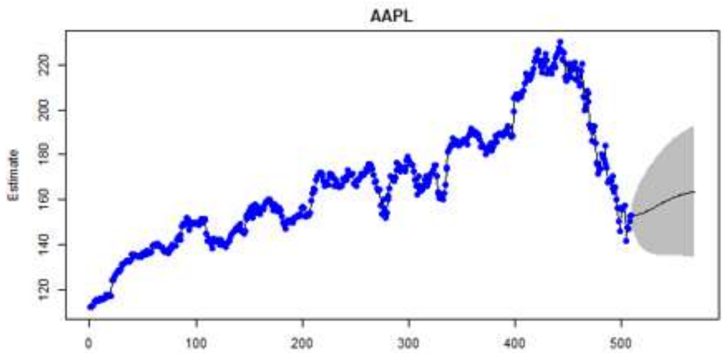
The ϕ term is a 6x6 matrix whose diagonal entries each represent the direct effect of the first lag on its respective company's current stock price. This interpretation is identical to that of an AR model. For instance, $\phi_{11} = 9.82\text{e} - 01$ (see figure 8 for coefficient values in appendix) corresponds to Apple. So Apple's stock price yesterday has a 98.2% effect on today's stock price.

The second coefficient $\phi_{1,2} = -\ 1.72\text{e} - 02$ corresponding to $x_{t-1,2}$, or the second time series adobe, means that Adobe's stock price yesterday has a -17.2% effect on Apple's stock price. The negative magnitude indicates Apple and Adobe stock price move in different directions; |-17.2%| < |100%| means that Apple reacts only partially to Adobe's stock price (or to only 17.2% of Adobe's movements). This is the highest coefficient among the other companies (excluding Apple itself), so Adobe influences Apple the most perhaps because their operations are intertwined, or perhaps of confounding variables.

The rest of the ϕ coefficients in figure 8 (in appendix) can be interpreted similarly relative to Apple. All the ϕ coefficients are < |1|, confirming the causality of our data and state space model.

Q is the covariance matrix for the w_t error term. Q.q11 = 6.73 corresponds to Apple's variance of error term, and is large because 1 < 6.73, meaning that Apple's stock price is influenced by a lot of randomness. This is true and confirms reality because stock movements are unpredictable. Q.q12 = 4.97 is the covariance between Apple and Adobe, so their error terms are strongly positively correlated. All the other Q terms can be interpreted similarly, all of which are greater than 1 meaning randomness is the key influencer.

Forecasting: The state space forecasted Apple's stock price estimate with an upward trend, and an increasing volatility with a rounded-out shape similar to the ARIMA model. The subtle distinction here though is that the rounded forecast area follows the shape of the trend more closely, as seen in the other series too in figure 7.2 (see appendix).



The MSE of this state space model will be in terms of prices, which ends up being a large ($) amount. The MSE would be more meaningful in terms of returns, especially when comparing to the other models. So I fit the state space model to diff(log(prices)) instead of just raw stock prices. I determined that the MSE in terms of returns is 0.0004094558, or .04094%. This means the state space model forecasts daily returns on average with a .0409% error rate.

The state space model does indeed help shed light on our projected insights. It forecasts stock trends/prices/returns, and volatilities so we know periods of higher risk. It also very easily helps analyze the effects of movements from other companies' stock prices/returns on Apple.

Model Comparison

Qualitatively:

The ARIMA model with drift assumes that the transformed data is stationary and roughly standard normal, which doesn't hold because of the clusters of volatility. Thus, the AR(1) + GARCH(1,1) model with a trend component better fits the data because it accounts for a non-constant standard deviation. The State Space model assumes the process starts with a normal vector, and also assumes the error term is standard normal. On the other hand, the GARCH model relaxes the assumption of normality, and is made to fit non-stationary times series.

The forecast of the GARCH is more profound than the ARIMA's. The ARIMA(2,1,2) forecasted prices, but its volatility, indicated by the wide round blue regions, was increasingly infinite. The AR + GARCH forecast forecasted σ_t so its confidence band areas shrank and expanded instead of constantly increasing. It accurately predicted the periods of more volatile fluctuations, which the ARIMA and State Space model failed to do. This provides analysts more information about apple's periods of volatility, helping investors better time and predict the market.

However, the State Space model's strength was that its coefficients were more comprehensive, which let us cross analyze the effects of different stocks conveniently. The AR coefficients revealed how the lag of a certain company along with the others' affected its current stock price. The covariance matrix of the error terms revealed how much randomness is inherent in a stock, and how its random error behaves with respect to those of the other companies. Additionally, state space model is helpful when dealing with incomplete data by smoothing and filling in the NA values. But in this stocks dataset, such a capability is not needed.

Overall, the **AR(1) + GARCH(1,1)** was the **best model for Stocks data** because it forecasts both returns and volatilities, and is well suited for dealing with non-stationary time series like stock movement.

Quantitatively:

Mean Square Error
$MSE_{ARIMA(2,1,2)} = .02525\%$
$MSE_{AR(1) + GARCH(1,1)} = .021837\%$
$MSE_{State\ Space\ Model} = .04094\%$.

Though all superficial, these small forecast errors over long periods of time or with large trade volume can compound into a huge deviation and adversely impact investing decisions. The AR(1) + GARCH(1,1) had the lowest MSE.

Though MSE shouldn't be the sole deciding factor, the GARCH model is still the best in modelling and forecasting Stocks data. In the GARCH model, 0% of the observations lay outside the yellow 95% confidence yellow region. Moreover, it forecasted the volatility σ_t (standard deviation), unlike the other models which gives investors more information and a different perspective than just expected returns.

Information Criterion
It would be worthwhile to see if combining ARIMA + GARCH model would improve results even more. So I combined the old ARIMA(2,1,2) model with the GARCH model. Because I am fitting data that is already transformed and differenced, I can remove the d=1 and just use ARMA(p=2, q=2).

Ultimately, I compared ARIMA(2,2) + GARCH(1,1) and, our original AR(1) + GARCH(1,1). To make a fair comparison, I fit an ARMA(2,2) + GARCH(1,1) under a Student's t distribution as well, with same package rugarch. Then, I compared information criterion, as shown below:

AR(1) + GARCH(1,1) with student's t distribution
Akaike -5.8916   Bayes -5.7695

ARMA(2,2) + GARCH(1,1) with student's t distribution
Akaike -5.8799   Bayes -5.7316

The order of AR(1) + GARCH(1,1) model is superior, supported by the lower AIC and BIC values. Another reason it is clearly better is that BIC favors parsimonious models by more heavily penalizing models with a greater number of parameters. ARMA(2,2) has 3 more parameters than AR(1), yet the AR(1)'s BIC is lower. One takeaway here is that even though a particular order (p=2,q=2) is optimal in a pure ARMA model, combining another model such as the GARCH(1,1) affects the original optimum.

Discussion

Throughout this report, I holistically analyzed the fits of different models and determined the best model was AR(1) + GARCH(1,1). Quantitative metrics, that I used, included Mean Square Error, information criterion (AIC/AICC/BIC), and confidence intervals. Qualitative aspects included the capabilities of each model such as its ability to deal with non-stationarity, its assumptions such as how strictly it relied on the normality assumption, and its insightfulness such as forecasting volatilities not just returns.

Having access to data on Apple's Operations can help identify exogenous variables in the state space model, which can improve its forecasting capabilities. For example, Iphone innovations hype up earnings announcements, and drive up stock price. Predicting its release and revenues from demand, and using this information to combine a linear regression with categorical and exogenous variables can add more layers of depth and improve forecast accuracy.