Kari Palmier
CSC 425 – Final Project

New York Stock Exchange Time Series Analysis Project

Introduction

- I was interested in how to utilize time series analysis to analyze and make predictions on a real world stock dataset
- The dataset I chose was the Kaggle New York Stock Exchange dataset from 1/4/2010 through 12/30/2016.
  - Kaggle dataset was for S&P 500 companies
  - Kaggle provided files with stock prices, fundamentals, and security information (company address, sector, etc)
  - Kaggle csv data file used had data that had been adjusted to account for any splits – used adjusted close prices
  - Used security information to select stocks to analyze based on sector (chose technology stocks)
- I selected several technology stocks to analyze before choosing a final one for modeling
  - Apple, Cisco, Alphabet Class C, Alphabet Class A, HP Inc., Intel Corp, Microsoft Corp, Nvidia Corp., Oracle Corp, Red Hat Inc., Texas Instruments, Western Digital, Xerox Corp, Yahoo Inc.

Exploratory Analysis

Exploratory Analysis

- Evaluated stocks at daily, weekly, and monthly granularity
  - Daily stocks had 1762 data points each
  - Weekly stocks had 587 data points each
  - Monthly stocks had 84 data points each
  - Generated weekly data by averaging value per week
    - Assigned date to Friday of week
- Evaluated normality using Jacque Bera test (p val > 0.05 = normal)
- Evaluated autocorrelation using Ljung Box test (p val > 0.05 = no autocorrelation)
- Evaluated stationarity using Augmented Dickey Fuller zero mean no trend, constant mean no trend, and constant mean with trend tests (p val < 0.05 = stationary series)
- Plotted time series data versus time, ACF, PACF, and normal probability plots
- Plotted CCF of each time series versus the rest

Exploratory Analysis

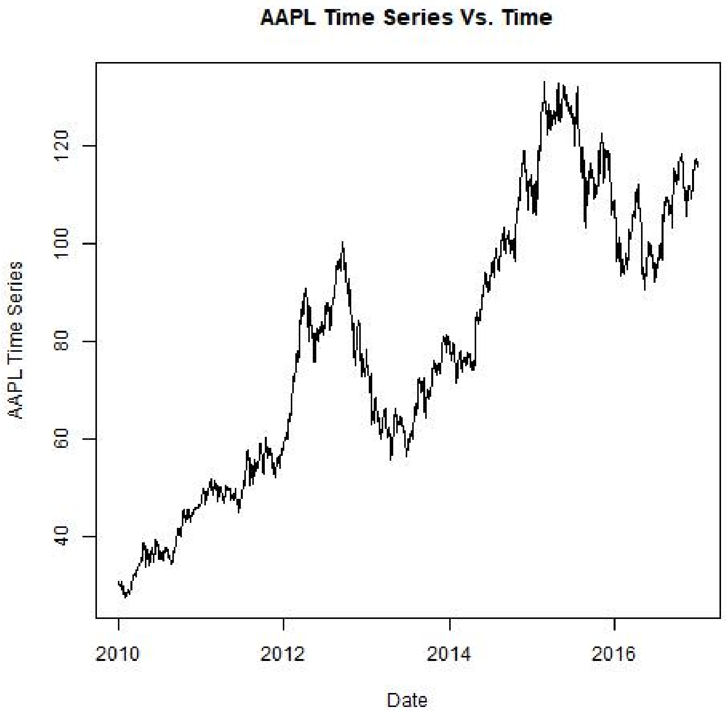
- Performed several transformations on stock adjusted close prices in order to evaluate stationarity and normality
  - Log transformation of adjusted close prices
    - Still exhibited a time trend for all granularities
  - Difference of adjusted close prices
    - Exhibited stationarity for most granularities and stocks
    - Only a half of the monthly granularities were normal
    - No daily or weekly were normal
  - Difference of log transformed adjusted close prices
    - Similar results as difference of adjusted close prices
  - Gross returns from adjusted close prices
    - Aug. Dickey Fuller showed non-stationary behavior for all granularities
  - Simple returns from adjusted close prices
    - Exhibited stationarity for most granularities and stocks
    - Most monthly and some weekly were normal

Exploratory Analysis

- Chose to use weekly granularity due to the number of points
  - 587 split into 80% training, 20% testing:
  - 470 training and 117 testing
  - Monthly only had 84 total points – not enough for a good split
- Chose to use the simple return time series transformation
  - Had most normal and stationary results for weekly granularity
- Chose to model Apple stock
  - Had normal behavior and was stationary at weekly granularity
  - Had well behaved ACF and PACF plots
  - Had limited autocorrelation (first lag only)
  - Had possibility of other stocks being used to predict it (from CCF)
  - Was an interesting stock to me to evaluate

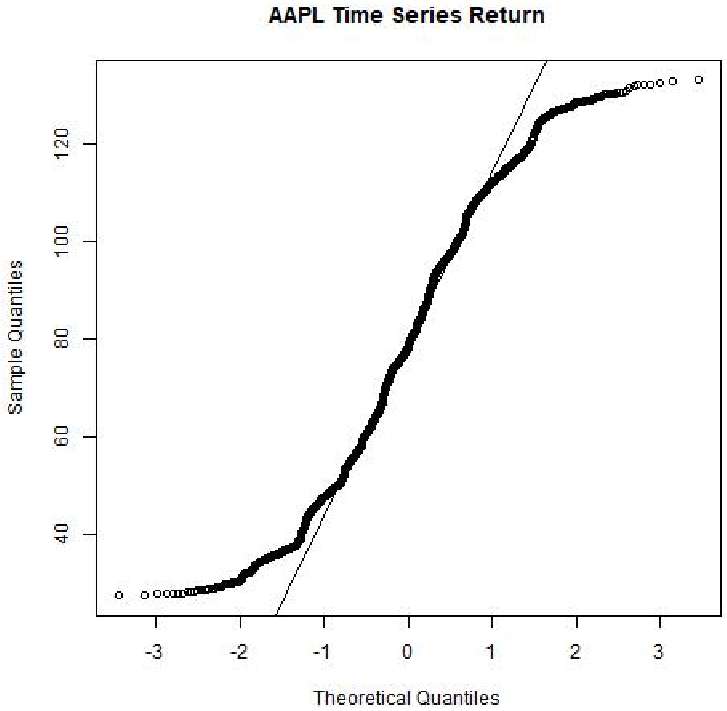
Exploratory Analysis

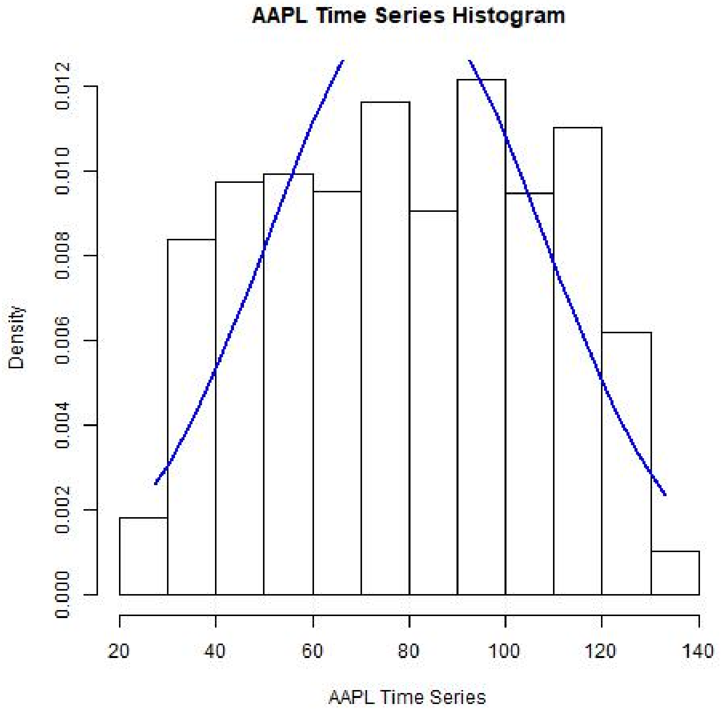
Apple Daily Closing Price Plots



Scatterplot vs. time is non-stationary with positive trend

Normal probability shows thick tails

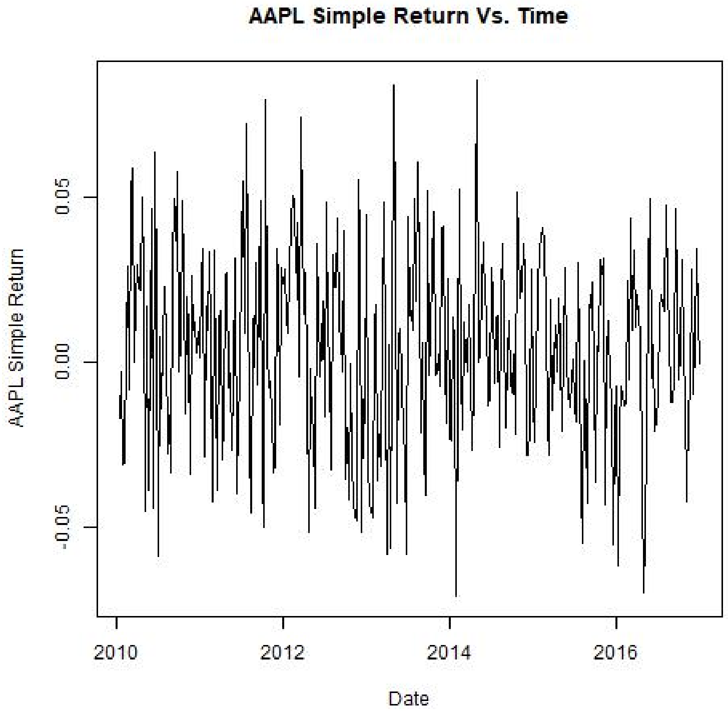




Histogram also shows thick tails

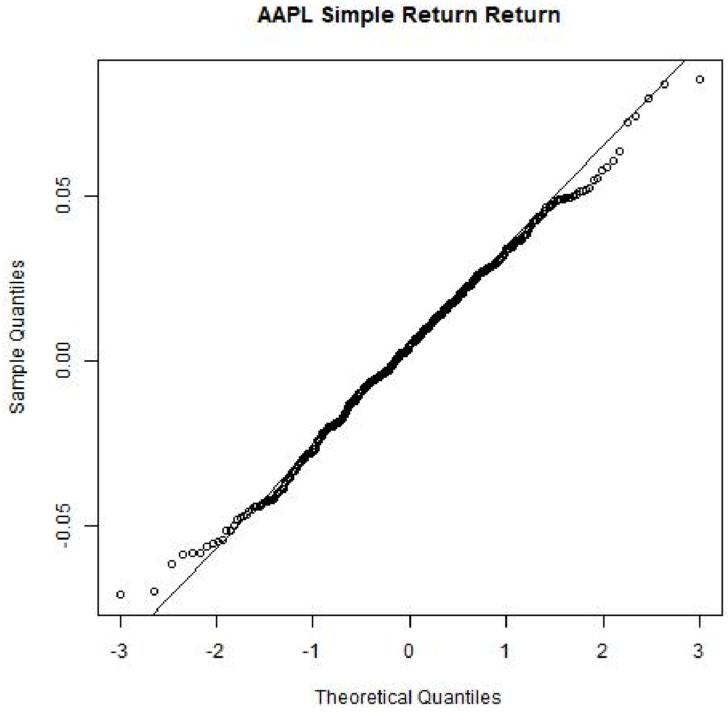
Exploratory Analysis

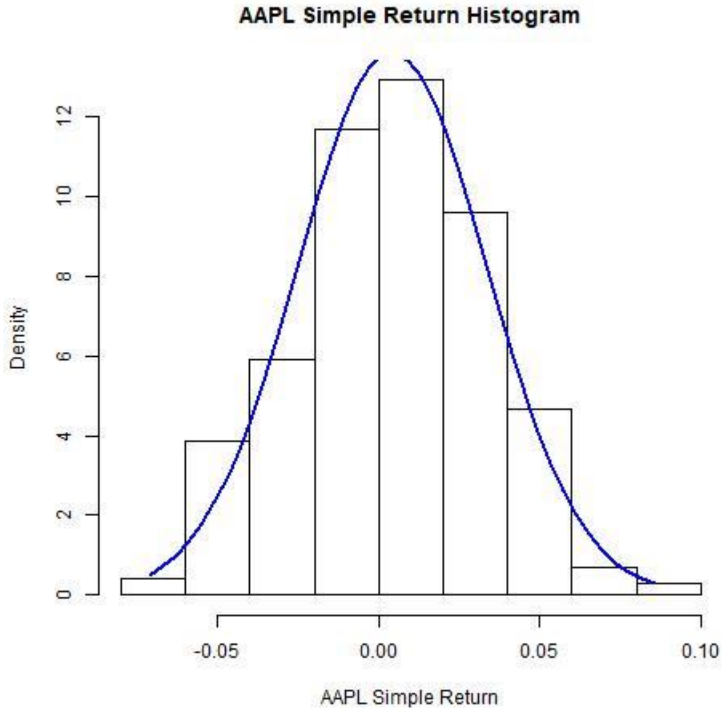
Apple Simple Return Plots



Scatterplot vs. time shows stationarity

Normal probability shows normality

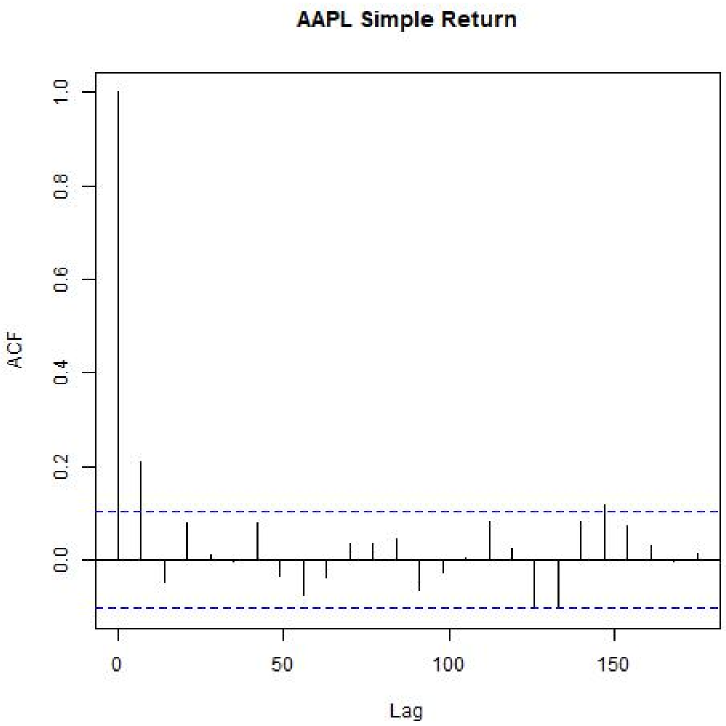


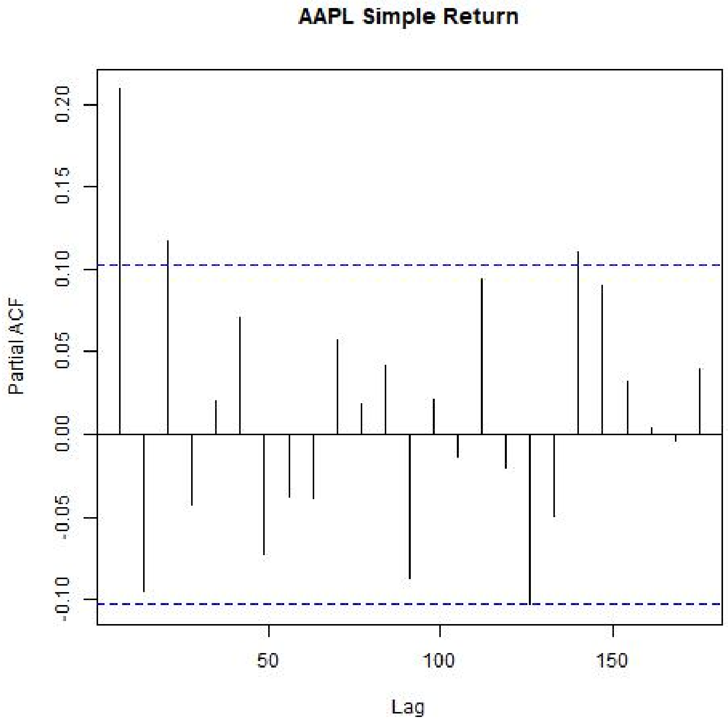


Histogram shows normal distribution

Exploratory Analysis

Apple Simple Return Plots





ACF shows 1 significant autocorrelation at lag 1

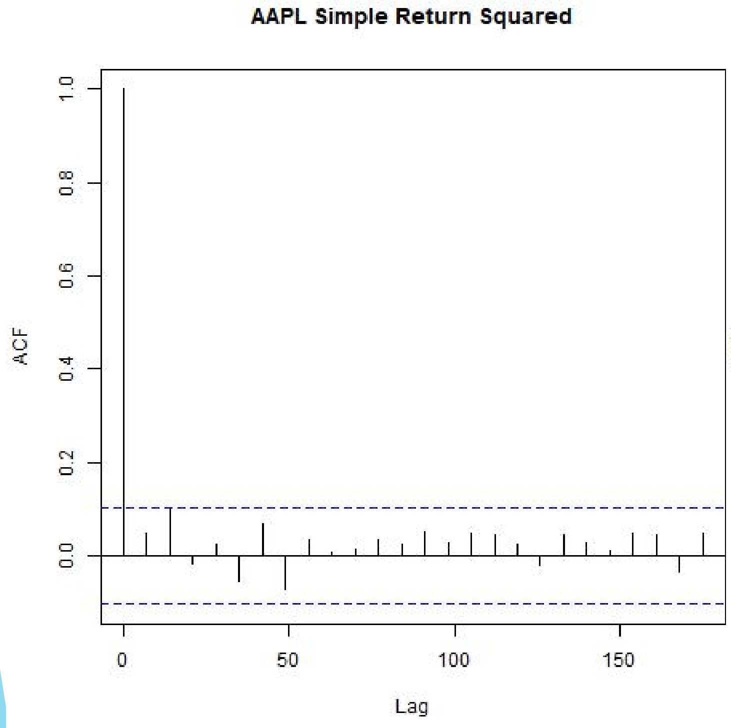
PACF a few lags over confidence limits
* Values are very low magnitude so could be noise

Exploratory Analysis

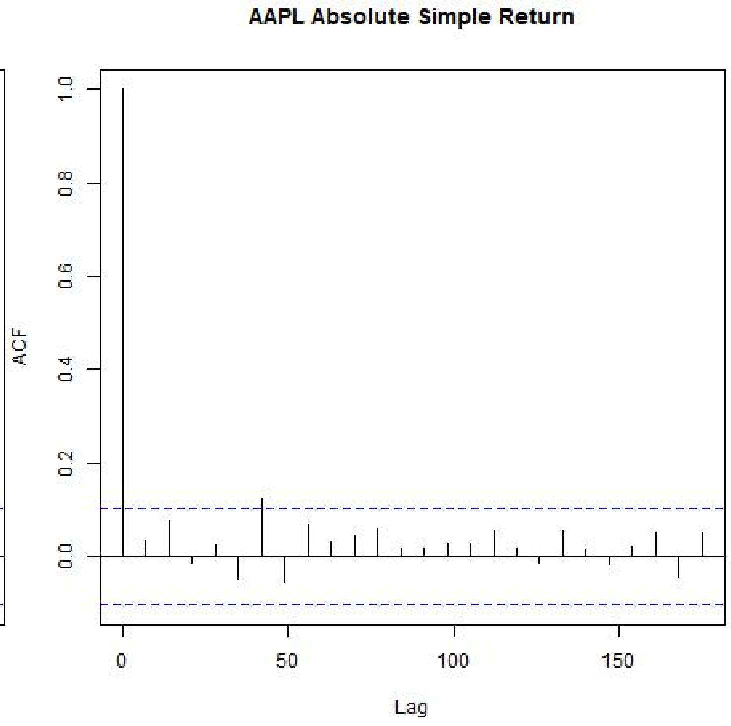
- Apple simple return time series test results

Number of Lags used in Ljung Box and ADF Tests	Weekly Simple Return JB Test P Value	Weekly Simple Return Kurtosis Test P Value	Weekly Simple Return Skewness Test P Value	Weekly Simple Return Ljung Box Test P Value	Weekly Simple Return No Intercept No Trend DF Test P Value	Weekly Simple Return Constant Intercept, No Trend DF Test P Value	Weekly Simple Return Constant Intercept With Trend DF Test P Value
5 Lags	0.5421	0.2734	0.6999	1.68E-03	0.01	0.01	0.01
10 Lags	0.5421	0.2734	0.6999	4.84E-03	0.01	0.01	0.01
25 Lags	0.5421	0.2734	0.6999	2.38E-03	0.01572	0.06898	0.2343

Exploratory Analysis



ACF of squared Apple simple returns shows no ARCH effect



ACF of absolute of Apple simple returns shows no ARCH effect

Number of Lags Used	Squared Weekly Simple Return Ljung Box Test P Value	Absolute of Weekly Simple Return Ljung Box Test P Value
5 Lags	0.3021	0.5932
10 Lags	0.4175	0.1939
25 Lags	0.8155	0.66

Ljung Box Tests indicate no ARCH effect (no autocorrelation)

Modeling

Modeling Procedure

- Ran auto.Arima on Apple simple return time series
  - Recommended MA(1) model
  - Recommendation matches ACF with lag 1 significance
- After auto.Arima, ran several different models using the following procedure
  - Created model using Arima function
  - Performed coefficient test using coeftest
  - Performed Jacque Bera normality test on model residuals
  - Generated the polynomial roots of the model
  - Performed Ljung Box test of residuals with the 5, 10, and 25 Lags
  - Performed Augmented Dickey Fuller zero mean no trend test, constant mean no trend test, and constant mean with trend tests
  - Plotted residuals versus time, normal probability plot of residuals, and ACF of residuals
  - Performed back testing with 80% training, 20% testing

Modeling Procedure

- Modeled MA(1), AR(1) and ARMA(1) to compare results
  - Had similar BIC values and back testing MAPE values
  - Jacque Bera test showed all residuals were normal
  - All Augmented Dickey Fuller tests showed residuals were all stationary
  - Ljung Box tests showed residuals were white noise for all lags (AR was not for lag of 25)
- Modeled MA(1) with X regressors of other stocks to see their impact
  - Used all additional time series at current time (no delay) from CCF plots
  - Initially added Cisco, Alphabet Class C, Alphabet Class A, Oracle Corp, Red Hat Inc., Texas Instruments, Yahoo Inc.
  - Eliminated all non-significant stocks one by one
  - Result was model with Alphabet Class C, Oracle Corp., and Texas Instruments
    - Residuals were not normal even though BIC values were lower than no regressors
  - Modeled with all combinations of final 3 stocks – none had normal residuals

Modeling Procedure

- Model results for all models

Model Name	Model Order	All Coefs Sig?	Model BIC	Residual JB Normal Test	Residual Ljung Box 5 Lag Test P Val	Residual Ljung Box 10 Lag Test P Val	Residual Ljung Box 25 Lag Test P Val	Residual No Interc No Trend DF Test P Value	Residual Const Interc No Trend DF Test P Value	Residual Const Interc W Trend DF Test P Value	Model Back Testing MAPE	Residual QQ Plot Tail Behavior	X Regressors
Apple_SimpleRtn_All_ARIMA_0_0_1_X_Regs	1	No	-1624.05	0.1802	0.2268	0.00038	0.00175	0.01	0.01	0.01	1.10053	Thick	CSCO, GOOG, GOOGL, ORCL, RHT, TXN, YHOO
Apple_SimpleRtn_ARIMA_0_0_1	1	Yes	-1541.1	0.8009	0.2332	0.2131	0.1382	0.01	0.01	0.01	1.22655	None	None
Apple_SimpleRtn_ARIMA_1_0_0	1	Yes	-1537.3	0.8532	0.06947	0.0827	0.04285	0.01	0.01	0.01	1.124128	None	None
Apple_SimpleRtn_ARIMA_1_0_1	2	Yes	-1538.4	0.7367	0.6798	0.3961	0.1991	0.01	0.01	0.01	1.205954	None	None
Apple_SimpleRtn_No_GOOGL_ARIMA_0_0_1_X_Regs	1	No	-1630	0.1805	0.2265	0.00038	0.00175	0.01	0.01	0.01	1.093953	Thick	CSCO, GOOG, ORCL, RHT, TXN, YHOO
Apple_SimpleRtn_No_GOOGL_RHT_ARIMA_0_0_1_X_Regs	1	No	-1635.8	0.1803	0.2315	0.00038	0.00158	0.01	0.01	0.01	1.092434	Thick	CSCO, GOOG, ORCL, TXN, YHOO
Apple_SimpleRtn_No_GOOGL_RHT_YHOO_ARIMA_0_0_1_X_Regs	1	No	-1641.2	0.1304	0.2472	0.00032	0.00107	0.01	0.01	0.01	1.056037	Thick	CSCO, GOOG, ORCL, TXN
Apple_SimpleRtn_W_GOOG_ARIMA_0_0_1_X_Regs	1	Yes	-1611.3	0.6467	0.0697	0.00131	0.00157	0.01	0.01	0.01	1.013589	Thick	GOOG
Apple_SimpleRtn_W_GOOG_ORCL_ARIMA_0_0_1_X_Regs	1	Yes	-1635.9	0.3465	0.2372	0.00123	0.00328	0.01	0.01	0.01	1.154316	Thickish	GOOG, ORCL
Apple_SimpleRtn_W_GOOG_ORCL_TXN_ARIMA_0_0_1_X_Regs	1	Yes	-1644.3	0.167	0.271	0.00055	0.00268	0.01	0.01	0.01	1.074363	Thick	GOOG, ORCL, TXN
Apple_SimpleRtn_W_GOOG_TXN_ARIMA_0_0_1_X_Regs	1	Yes	-1638.2	0.1603	0.1435	0.00033	0.00192	0.01	0.01	0.01	0.985778	Thick	GOOG, TXN
Apple_SimpleRtn_W_ORCL_ARIMA_0_0_1_X_Regs	1	Yes	-1603.2	0.5477	0.5438	0.03175	0.04749	0.01	0.01	0.01	1.295805	Thin	ORCL
Apple_SimpleRtn_W_TXN_ARIMA_0_0_1_X_Regs	1	Yes	-1608.8	0.3794	0.2543	0.00438	0.01671	0.01	0.01	0.01	1.098201	Thickish	TXN
Apple_SimpleRtn_W_TXN_ORCL_ARIMA_0_0_1_X_Regs	1	Yes	-1624.3	0.3557	0.5025	0.00522	0.01566	0.01	0.01	0.01	1.158269	Thickish	ORCL, TXN

Modeling Procedure

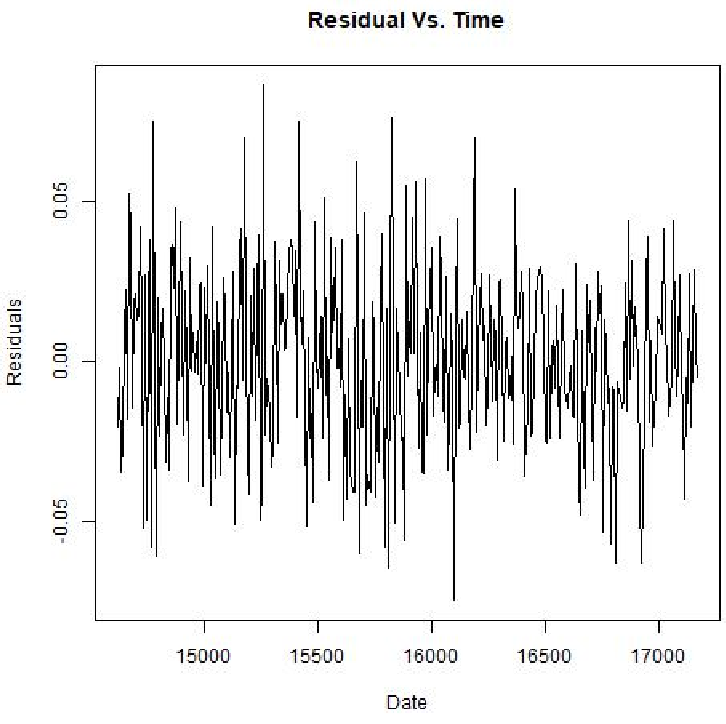
- MA(1), AR(1), and ARMA(1) model results for Apple simple return series

Model Name	Model Order	Model BIC	Model AR Poly Roots	Residual JB Normal Test	Residual Ljung Box 5 Lag Test P Val	Residual Ljung Box 10 Lag Test P Val	Residual Ljung Box 25 Lag Test P Val	Residual No Intercept, No Trend DF Test P Value	Residual Constant Intercept, No Trend DF Test P Value	Residual Constant Intercept With Trend DF Test P Value	Model Back Testing RMSE	Model Back Testing MAPE
Apple Simple Return MA(1)	1	-1541.1		0.8009	0.2332	0.2131	0.1382	0.01	0.01	0.01	0.02457	1.22655
Apple Simple Return AR(1)	1	-1537.3	-4.7888	0.8532	0.06947	0.0827	0.04285	0.01	0.01	0.01	0.02467	1.124128
Apple Simple Return ARMA(1)	2	-1538.4	3.4426	0.7367	0.6798	0.3961	0.1991	0.01	0.01	0.01	0.02489	1.205954

- Forecasted results with horizon set to 10 for all three models
  - MA(1) converged to the mean after 1 step ahead as expected
  - AR(1) and ARMA(1) essentially converged to the mean after 1 step ahead as well (were within 0.0002 of mean after 1 step ahead)
  - Back testing MAPE values were all within 0.1 of each other (negligible)
- Due to similarity in results, chose to use MA(1) model for final model
  - Probability plot showed smallest tails (they were all close)

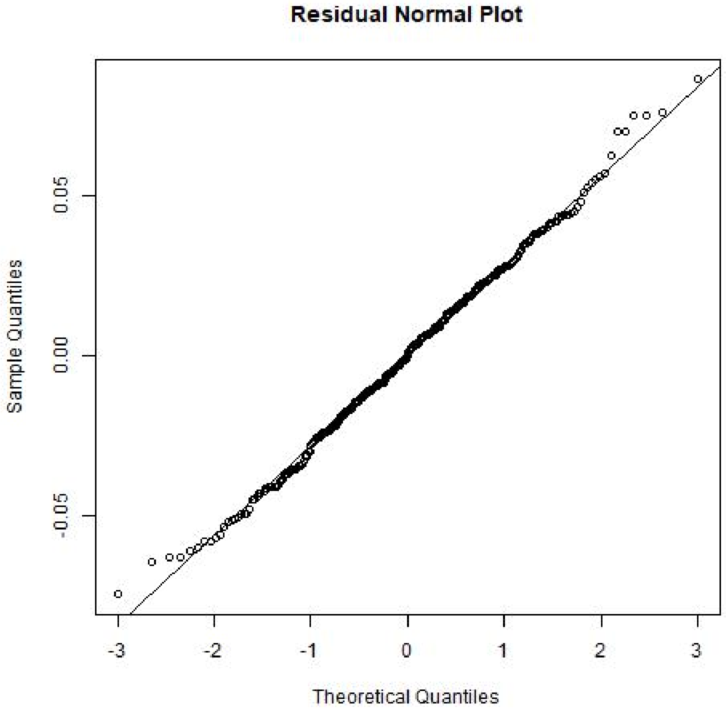
Modeling Procedure

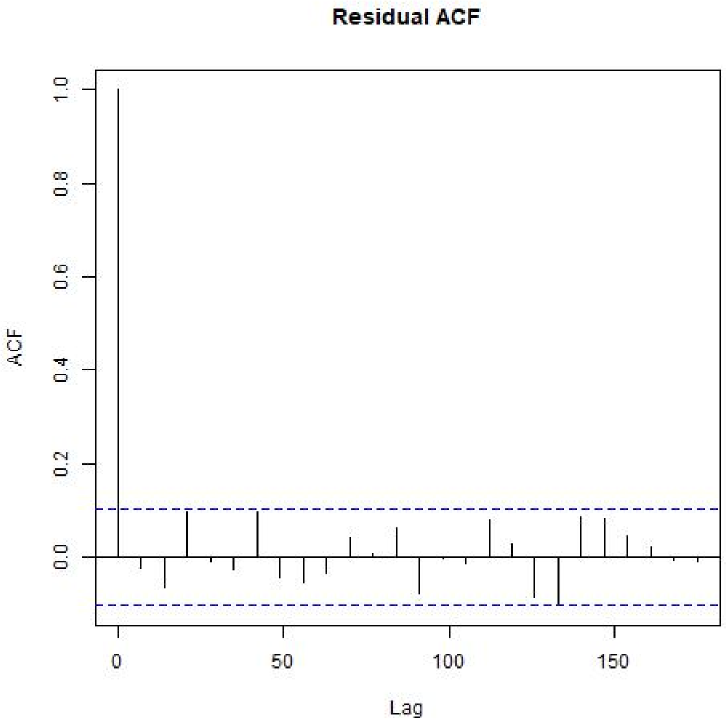
- MA(1) Model Residual Plots



Model residuals vs. time shows stationarity

Residual normal probability shows normality

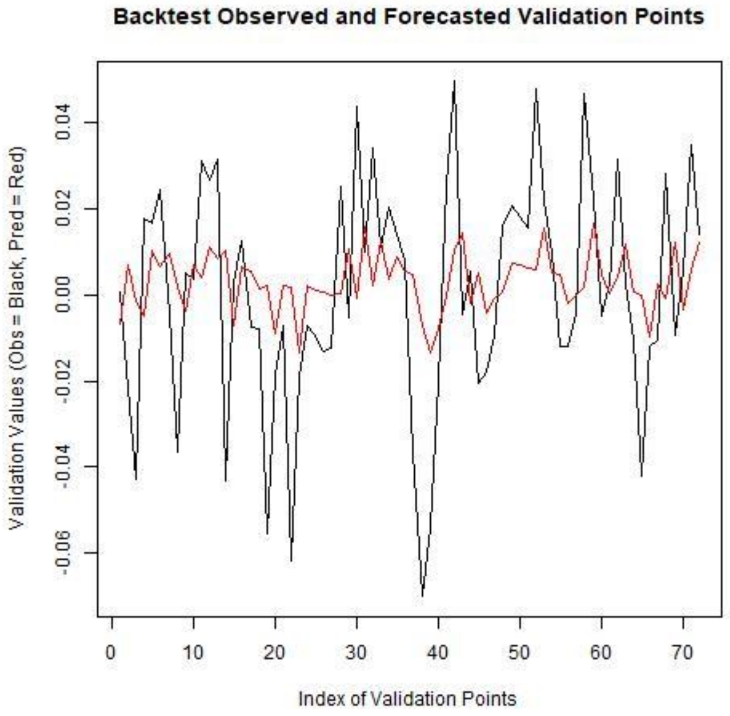




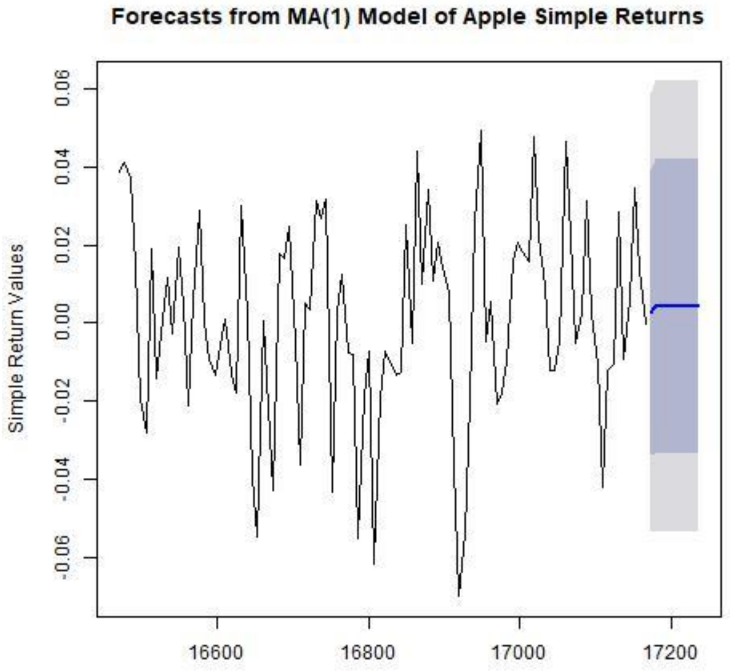
ACF shows no autocorrelation

Modeling Procedure

- MA(1) Model Forecasting Plots



Back testing observed values and forecasted values



Apple simple returns last 100 points with 10 forecasted values

Conclusions

- MA(1) model best captured behavior of Apple simple returns
- MA(1) model was found to be adequate for predictions
  - MA(1) is stationary and residuals passed all required tests
- 1-step ahead forecast shows increase in future week
  - Recommendation would be to buy the stock
- MA(1) model equation:

  $X_t = 0.0041 + a_t - 0.2634 * a_{t-1}$

- Limitations of study:
  - Model can only be used to forecast 1 step ahead due to model order of 1
  - Model only apply to Apple stocks from 1/4/2010 to 1/6/2017 (one week after end of data – one step ahead forecast)
  - Data was averaged over each week
    - Variation between days was removed by taking average