Statistical Modelling of the Capital Asset Pricing Model (CAPM)

Silvi Qemo[1] & Eahab Elsaid[1]

[1] Odette School of Business, University of Windsor, Canada

Correspondence: Eahab Elsaid, Odette School of Business, University of Windsor, Canada

Abstract

The purpose of this study is to derive a multiple linear regression model of the CAPM. More specifically, to test for other potential explanatory variables that can be added to the basic linear regression model for the expected returns on Apple Inc. The following explanatory variables were examined: share volume, outstanding shares, closing bid/ask spread, high/low spread and average spread. Using daily returns of Apple Inc. stock from 2007 till 2014 we were able to create a multiple linear regression model of CAPM that increase the R2 value from the basic linear regression model and enhances the amount of variability in the returns on an asset. This is an important modification that can help better forecast returns on assets.

Keywords: CAPM, multiple linear regression model, average spread, variability in the returns

1. Introduction

The Capital Asset Pricing Model (CAPM) is a theory credited to Sharpe (1964) and Lintner (1965) and was grounded on the work of Markowitz (1952, 1959), which dealt with portfolio theory and portfolio diversification theory (Fama and French, 2004). It explained the relationship between market risk and the expected return on a particular asset. An in depth analysis of the basic linear regression CAPM is provided in Supplemental File 1.

The purpose of this study is to analyze the CAPM and to try and move from the basic linear regression model of the CAPM to a multiple linear regression model of the CAPM. The response variable and explanatory variable in the basic linear regression model, CAPM, is the return on an asset and the return on the market respectively. Following the mathematical derivation of CAPM (Supplemental File 1), we further derive the model for a specific stock (Supplemental File 2), Apple Inc., and attempt to find other explanatory variables that may be added to a multiple linear regression model of the CAPM. We then proceed to check for collinearity and include interaction terms in the multiple linear regression model as well as conduct a residual analysis on the multiple linear regression model. We conclude with robustness testing (Supplemental File 3), in which other stocks are used during the same time period and sectional robustness testing by reducing the amount of data points tested for Apple Inc. stock.

The paper makes a contribution through the creation of a multiple linear regression model, (Equation 4), which we show to be statistically significant for Apple Inc. stock and further on in the robustness testing. The multiple linear regression model of the CAPM produced in this paper increases the multiple R^2 value from the basic linear regression model of the CAPM. Hence, enhancing the amount of variability in the returns on an asset that can be explained as compared to the basic linear regression model of the CAPM as is now. This is an important modification that can better help forecast returns on an asset.

2. Hypothesis Development

We test for other potential explanatory variables that can be added to the basic linear regression model of the CAPM for the expected returns on Apple Inc. (other than the S&P 500 returns which is already an explanatory variable in the basic model as the market returns in the CAPM). In this multiple linear regression model, the goal is to find other relationships that exist between the expected stock returns and other variables. The following is a list of the explanatory variables tested in the regression model:

- *Share Volume* – The share volume for daily files is the total number of shares sold on that day (CRSP, 2017a). A log transform was applied, taking the LN of VOL, as this does not change the distribution but helps to better graph the relationship with the monthly Apple Inc. returns (which are in decimals).

- *Outstanding Shares* – "*The unadjusted number of publicly held shares...recorded in 1000s*" (CRSP, 2017b). A log transform was applied, taking the LN of SHARES, as this does not change the distribution but helps to better graph the relationship with the monthly Apple Inc. returns (which are in decimals).

- *The Closing Bid/Ask Spread* – The spread "*is the difference between the closing bid and ask quotes for a security*" (CRSP, 2017c). Daily data was used.

- *The High/Low Spread* – Is the spread calculated as the difference between the highest ask price and the lowest bid price for that day. This is daily data rather than monthly data.

- *The Average Spread* – Is the spread calculated as the difference between the highest ask price and the lowest bid price for that day divided by the price of the stock at the end of the day. This is daily data rather than monthly data.

3. Data Collection & Methodology

When empirically testing CAPM, we chose to use Apple Inc. stock returns availability of data. The market returns used in this study are the S&P 500 market returns. All data was extracted from the Wharton Research Data Services (WRDS) of University of Pennsylvania under CRSP stock/security files, using both monthly and daily stock files.

For the basic linear regression model (Supplemental File 1), the observed values are the historical monthly returns on Apple Inc. and on the S&P 500 from January 2007 until December 2015. For the multiple linear regression model, we use the daily stock returns and daily market (S&P 500) returns. We use the daily returns in order to accommodate the spread variables which are daily values and not monthly values. Specifically, daily data was collected for the time period from January 2007 until June 6, 2014 instead of until the end of 2015 because there was a stock split that occurred after that day that affected the number of outstanding shares and the share price. This would have created a discrepancy if the subsequent data was also used in the regression model.

All statistical analysis was conducted using the program R (version 3.1.2 64 bit). The authors used a confidence interval of 95% or equivalently an alpha = 0.05. A basic linear regression test was used to show the relationship between monthly returns on Apple Inc. stock and monthly S&P 500 returns. Box plots, histograms and normal Q-Q plots were produced for each of these two variables as well (Supplemental File 2).

4. Results

We examine a multiple linear regression model of CAPM. Our purpose is to test for other explanatory variables that can be added to the basic linear regression model for the expected returns on Apple Inc. In this multiple linear regression model, the goal is to find other relationships that exist between the expected stock returns and other explanatory variables such as: share volume, outstanding shares, closing bid/ask spread, high/low spread and average spread. The two methods of regression used in the paper include backwards stepwise regression and forward selection regression. Stepwise regression looks to find explanatory variables that can be added or deleted from a model by searching through combinations of the explanatory variables. Backwards stepwise regression takes in the set of all the explanatory variables to begin with, in the model, and eliminates variables with the highest AIC (Akaike Information Criterion) in each step until only those variables that prove to be statistically significant remain in the model. Forward Selection Regression, on the other hand, starts with an empty model and adds explanatory variables that are most statistically significant first and continues to do so until there are no more statistically significant variables to be added into the model (Frees, 2010).

Table 1. Backwards Stepwise Regression Results

Table 1 Panel A

	Df	Sum of Sq	RSS	AIC
<none>			0.57501	-15113
lnshares	1	0.00134	0.57635	-15110
lnvol	1	0.00426	0.57927	-15101
avgspread	1	0.01073	0.58574	-15080
returnSP	1	0.32741	0.90242	-14272

Table 1 Panel A reports the results of the stepwise regression with the following explanatory variables for Apple Inc. daily stock data: logarithmic transform of share volume (lnshares), logarithmic transform of outstanding shares (lnvol), the closing bid/ask spread, the high/low spread, the average spread (avgspread) and the daily return on the

S&P 500 (returnSP). The backwards stepwise regression takes in all variables for evaluation to be eliminated in that step and a variable is only eliminated if eliminating that variable reduces the Akaike Information Criterion (AIC). This is repeated until no other variable can be eliminated. The first column represents the variables left after the stepwise regression was completed. The second column represents the degrees of freedom. The third column represents the sum of squares. The fourth column represents the residual sum of squares. The fifth column represents the Akaike Information Criterion (AIC).

Table 1. Panel B

Variable	Estimate
(Intercept)	0.391158
lnshares	-0.033246
lnvol	0.004198
avgspread	-0.186518
returnSP	0.926339

Table 1 Panel B reports the coefficient estimates of the statistically significant variables that the backwards stepwise regression test produced in Table 1 Panel A. The first column represents the variables that were indicated as being necessary in the multiple regression model by the stepwise regression method. The second column represents the coefficient estimates calculated by the stepwise regression test for each of the statistically significant explanatory variables in the multiple regression model. Producing the following multiple regression model:

$$E(r_A) = 0.391158 - 0.033246(lnshares) + 0.004198(lnvol) - 0.186518(avgspread) + 0.926339E(r_M).$$

Table 1 Panel A reports the variables that are statistically significant in the multiple linear regression model (with the returns on the Apple Inc. stock as the response variable): outstanding shares (ln transform), volume of shares (ln transform), average spread and the returns on the S&P (which is already established as the main explanatory variable in the basic linear regression model). Table 1 Panel B also provides us with a multiple linear regression line as follows:

$$E(r_A) = 0.39115 - 0.03326(lnshares) + 0.004198(lnvol) - 0.186518(avgspread) + 0.926339E(r_M) \quad (1)$$

Where the expected returns on the asset,$E(r_A)$, are the Apple Inc. monthly returns and the expected returns on the market,$E(r_M)$, are the S&P 500 monthly returns. All four explanatory variables were shown to be highly statistically significant

Table 2. Forward Selection Regression Results

	Estimate	Standard Error	t value	p value
(Intercept)	-0.0276192	0.0113051	-2.443	0.0146
lnshares	-0.0031875	0.0007203	-4.425	1.01e-05***
lnvol	0.0045788	0.0009698	4.722	2.48e-06***
avgspread	-0.1809787	0.0285367	-6.342	2.73e-10***
returnSP	0.9387729	0.0260504	36.037	< 2e-16***

Significance codes:　<0.001 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' '

Table 2 reports the coefficients of the multiple linear regression through a forward selection method with the significant variables produced by the backwards stepwise regression method. The estimate column indicates the coefficients calculated by the regression test. The standard error column represents a measure of variability in the coefficients calculated by the regression test. Small standard errors represents a good regression test meaning variability is low. The t-value column represents whether the coefficients of the variables are statistically significant in the model. The p value column represents the probability that the variable in the model is not significant. The reported p-values are low, which indicate that the variables are very significant.

Table 2 is a forward selection regression to test the significance of the variables in the backwards stepwise regression. Even though the estimates of the coefficients vary between the two models, both regression methods demonstrate statistically significant explanatory variables which is the main hypothesis test.

Table 3. Correlation Matrix

	returnsapple	returnSP	lnshares	lnvol	avgspread
returnsapple	1				
returnSP	0.60902487	1			
lnshares	-0.02929778	0.01782082	1		
lnvol	-0.04492766	-0.0845975	-0.3495068	1	
avgspread	-0.15310253	-0.1341192	-0.2309337	0.70680407	1

Table 3 reports the correlations between the explanatory variables which were found to be statistically significant in both Table 1 Panel B and Table 2 and the response variable. Returnsapple represents the daily returns on Apple Inc. stock (the response variable in the model). ReturnSP represents the daily returns on the S&P 500. Lnshares represents the logarithmic transform of the daily outstanding shares of Apple Inc. stock. Lnvol represents the logarithmic transform of the daily volume of shares of Apple Inc. stock. Avgspread represents the daily average spread data points of Apple Inc. stock. All daily data points are from the time period January 2007 until June 6, 2014.

Since the correlation matrix represented in Table 3 only captures relationships between two variables at a time, it is necessary to check for collinearity between the explanatory variables.

Table 4. VIF Results of the Explanatory Variables

Explanatory Variable	VIF
lnshares	1.13
lnvol	2.15
avgspread	2.02
returnSP	1.01

Table 4 reports the variance inflation factor (VIF) of each of the explanatory variables. The VIF is used to check collinearity between the explanatory variables.

In order to check collinearity it is important to look at the variance inflation factor (VIF) of each variable. Table 4 represents the VIF of each explanatory variable. As a rule of thumb, VIF should be less than 10 otherwise there would be major collinearity. As we see from these values, in Table 4, each VIF is significantly smaller than 10.

Interaction terms of the variables were then included in the multiple linear regression model in order to incorporate any relationships between any two variables in the model for a more accurate model. Table 5 includes the results from the backwards stepwise regression method with the interaction terms included in the model.

Table 5. Backwards Stepwise Regression including Interaction Terms Results

Table 5 Panel A

	Df	Sum of Sq	RSS	AIC
<none>			0.56840	-15128
lnshares x avgspread	1	0.001029	0.56943	-15127
lnshares	1	0.001044	0.56944	-15127
lnshares x lnvol	1	0.001127	0.56953	-15127
lnvol	1	0.001139	0.56954	-15127
lnvol x avgspread	1	0.001261	0.56966	-15126
returnSP x avgspread	1	0.003529	0.57193	-15119
returnSP	1	0.074843	0.64324	-14899

Table 5 Panel A reports the stepwise regression results with several explanatory variables and interaction terms included. All daily data points are from the time period January 2007 until June 6, 2014. The backwards stepwise regression takes in all variables for evaluation to be eliminated in that step and a variable is only eliminated if eliminating that variable reduces the Akaike Information Criterion (AIC). This is repeated until no other variable can be eliminated. The first column represents the variables left after the stepwise regression was completed. The second

column represents the degrees of freedom. The third column represents the sum of squares. The fourth column represents the residual sum of squares. The fifth column represents the Akaike Information Criterion (AIC).

Table 5 Panel B

Variable	Estimate
(Intercept)	-13.59205
lnshares x avgspread	0.10827
lnshares	0.98626
lnshares x lnvol	-0.06102
lnvol	0.84092
lnvol x avgspread	-0.09253
returnSP x avgspread	2.85782
returnSP	0.78544

Table 5 Panel B reports the coefficient estimates of the statistically significant variables that the backwards stepwise regression test produced in Table 5 Panel A. The first column represents the variables that were indicated as being necessary in the multiple regression model by the stepwise regression method. All daily data points are from the time period January 2007 until June 6, 2014. The second column represents the coefficient estimates calculated by the stepwise regression test for each of the statistically significant explanatory variables in the multiple regression model.

Table 6 represents the results of the forward selection regression method (which examines another potential multiple linear regression model) and it includes the explanatory variables that were shown to be necessary in the stepwise regression as well as all the individual variables needed for the interaction terms. The return on the S&P 500 remains as the most statistically significant explanatory variable. In addition, the interaction terms of lnshares x lnvol (the interaction term between share volume and outstanding shares) and returnSP x avgspread (the interaction term between the S&P 500 returns and the average spread) are statistically significant as well with their respective t-values being larger than 1.966 or equivalently their respective p-values being smaller than 0.05.

Table 6. Forward Selection Regression including Interaction Terms Results

Table 6 Panel A

	Estimate	Standard Error	t value	p value
(Intercept)	-20.00155	9.22223	-2.169	0.030221*
lnshares	1.45439	0.67296	2.161	0.030809 *
lnvol	1.25689	0.56786	2.213	0.026993 *
returnSP	0.77901	0.05047	15.434	< 2e-16 ***
avgspread	-26.10000	22.88161	-1.141	0.254161
returnSP x avgspread	2.97912	0.84720	3.516	0.000448 ***
lnshares x lnvol	-0.09139	0.04144	-2.205	0.027545 *
lnshares x avgspread	1.98214	1.64386	1.206	0.228055
lnvol x avgspread	-0.06864	0.05010	-1.370	0.170814

Multiple R^2: 0.3916, Adjusted R^2: 0.389

Significance codes: <0.001 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' '

Table 6 Panel A includes the explanatory variables that were shown to be statistically significant in the backwards stepwise regression method along with the corresponding explanatory variables needed for the statistically significant interaction terms in the backwards stepwise regression model.

Table 6 Panel B

	Estimate	Standard Error	t value	p value
(Intercept)	-9.12020	6.97314	-1.308	0.191
lnshares	0.66152	0.50880	1.300	0.194
lnvol	0.57028	0.41434	1.376	0.169
returnSP	0.76532	0.04922	15.551	< 2e-16 ***
avgspread	-0.15766	0.03232	-4.879	1.16e-06 ***
returns x avgspread	3.26076	0.81749	3.989	6.90e-05 ***
lnshares x lnvol	-0.04135	0.03023	-1.368	0.172

Multiple R^2: 0.3898, Adjusted R^2: 0.3879

Significance codes: <0.001 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' '

Table 6 Panel B includes only the statistically significant explanatory variables found in Table 6 Panel A along with the explanatory variables needed for the statistically significant interaction terms in Table 6 Panel A.

Table 6 Panel C

	Estimate	Standard Error	t value	p value
(Intercept)	0.0030865	0.0007103	4.345	1.47e-05 ***
returnSP	0.7479486	0.0491298	15.224	< 2e-16 ***
avgspread	-0.0710615	0.0228075	-3.116	0.00186 **
returns x avgspread	3.6675923	0.8124938	4.514	6.76e-06 ***

Multiple R^2: 0.3828, Adjusted R^2: 0.3819

Significance codes: <0.001 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' '

Table 6 Panel C includes only the statistically significant explanatory variables found in the Table 6 Panel B along with the explanatory variables needed for the statistically significant interaction terms in Table 6 Panel B.

All daily data points are from the time period January 2007 until June 6, 2014. Additionally the intercept is the y-intercept of the model. Table 6 has four columns. The estimate column represents the coefficients calculated by the regression test. The standard error column is a measure of variability in the coefficients calculated by the regression test. Small standard errors represents a good regression test, meaning variability is low. The t-value column represents whether the coefficients of the variables are statistically significant in the model. The p-value column represents the probability that the variable in the model is not significant. At the bottom of each panel, the multiple R^2 represents the amount of variability in the response variable (daily returns on Apple Inc. stock) that is explained by that particular model.

Two possible multiple linear regression models were provided. The first model, Table 6 Panel A, is the result of the backwards stepwise regression method. The second model, Table 6 Panel C, is the result of the forward selection regression test. The first model produced is shown as Equation 2 and the second model produced is shown as Equation 3.

$$E(r_A) = -20.00155 + 1.45439(lnshares) + 1.25689(lnvol) - 26.1(avgspread) + 0.77901(returnSP) + 2.97912(returnSP \ x \ avgspread) - 0.09139(lnshares \ x \ lnvol) + 1.98214(lnshares \ x \ avgspread) - 0.06864(lnvol \ x \ avgspread) \tag{2}$$

$$E(r_A) = 0.0030865 + 0.7479486(returnSP) - 0.0710615(avgspread) + 3.6675923(returnSP \ x \ avgspread) \tag{3}$$

Where the following is the explanation of each variable:

- lnshares is the logarithmic transform of the daily outstanding shares of Apple Inc. stock.

- lnvol is the logarithmic transform of the daily volume of shares of Apple Inc. stock.

- returnSP is the daily return on the S&P 500. Also known as $E(r_M)$ in the basic linear regression model, the expected return on the market.

- avgspread is the daily average spread of Apple Inc. stock.

- returnSP x avgspread is the interaction term of returnSP and avgspread.

- lnshares x ln vol is the interaction term between lnshares and lnvol.

- lnshares x avgspread is the interaction term between lnshares and avgspread.

- lnvol x avgspread is the interaction term between lnvol and avgspread.

From the above modification of the stepwise regression results to the forward selection regression results (Equation 2 and Equation 3), there is a decrease in the coefficient of determination (R^2) from 0.3916 to 0.3898 to 0.3828. However, Equation 3, establishes a clear significance for the need of the interaction term between average spread and returnSP.

A residual analysis of both the first model, Table 6 Panel A or Equation 2, and the second model, Table 6 Panel C or Equation 3, was conducted. Refer to Supplemental File 3 for these results. The residual plots reported in Supplemental File 3, show a random distribution for the residuals which indicates that the assumption that the residuals are independent, random variables holds. Additionally, the QQ plots reported in Supplemental File 3 show relatively straight lines for both models, which implies that the residuals are normally distributed as they should be. Since the residuals are normally distributed this indicates that the mean of the values $(y - \hat{y})$ are close to zero (as a standard normal distribution has a mean of zero). This indicates that the actual and predicted values of the dependent variable are close, hence, the regression line is a good representation of the data set in both models. However, the Robustness Testing section below shows that Equation 3 is the only one out of the two multiple linear regression models that can be generalized to other stocks and sectional data. Therefore, a plausible multiple linear regression model of the CAPM is as follows:

$$E(r_A) = \beta_0 + \beta_1(returnSP) + \beta_2(avgspread) + \beta_3(avgspread \; x \; returnSP) \tag{4}$$

4.1 Robustness Testing

Two different types of robustness tests were conducted. The first was a sectional robustness test, where a smaller sample of the data points from the Apple Inc. stock was used. Twenty days and their respective daily variables were extracted from the beginning of each year from 2007-2014, a total of 160 data points. Table 7 represents the result of the sectional robustness test and the multiple linear regression model formed in Equation 4 applied to this smaller sample of data to examine if this model still increases the multiple R^2 value and if the extra explanatory variables are still statistically significant.

Table 7. Sectional Robustness Testing for Apple Inc. Stock

	Estimate	**Standard Error**	**t value**	**p value**
(Intercept)	0.005993	0.003597	1.667	0.09764 .
returnSP	0.931905	0.378357	2.463	0.01486 *
avgspread	-0.297102	0.106543	-2.789	0.00595 **
returns x avgspread	-0.741305	8.325725	-0.089	0.92917

Multiple R^2: 0.2127, Adjusted R^2: 0.1975

Significance codes: <0.001 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' '

Table 7 represents the results of the robustness tests using twenty daily data points from the beginning of each year from 2007 until 2014, a total of 160 data points. ReturnSP is the daily return on the S&P 500. Avgspread is the daily average spread of Apple Inc. stock. ReturnSP x avgspread is the interaction term of returnSP and avgspread.

In addition, the basic linear regression model (CAPM) produced for this specific sample of 160 daily points of Apple Inc. stock created a model with a multiple R^2 value of 0.1734 and an adjusted R^2 value of 0.1682 (refer to the results reported in Supplemental File 3). Hence the multiple linear regression model above in Table 7 increased the multiple R^2 value by 0.0393. This is an almost 4% increase in the multiple R^2 value. This means that an increased amount of variability (almost 4%) in the response variable (daily returns on Apple Inc. stock) can be explained through this new multiple linear regression model.

The second robustness test performed included examining other stocks to test whether the multiple linear regression model, Equation 4, holds and can be generalized. First, a list of all constituents of the S&P 500 was extracted. Then ten stocks from this list were randomly chosen using Microsoft Excel and its random function, excluding Apple Inc. stock. The stocks that were chosen had to have daily data available from January 2007 to June 6, 2014. This time frame was chosen in order to match with the data time frame used for Apple Inc. stock previously. All of the randomly selected ten stocks showed significance in at least one of the added explanatory variables, avgspread or the interaction term (avgspread x returnSP). Eight out of the ten stocks showed high statistical significance in the interaction term (avgspread x returnSP). The other two stocks showed statistical significance in the avgspread term but no significance in the interaction term, (avgspread x returnSP). However, including the interaction term is necessary as shown by the majority of the stocks. Therefore, all ten stocks showed that the multiple regression model that was produced in this study improved the multiple R^2 value from the basic linear regression model that is the CAPM and showed significance in either or both variables in the multiple linear regression model produced. Additionally, contrary to the results found in Fama and French (2004), the importance of CAPM is shown once again in these generalized regression tests as all regression tests clearly indicate the market returns variable as the most statistically significant variable in all the models. All regression tests run for this part of the robustness testing are provided in Supplemental File 3.

5. Conclusion

The multiple linear regression model, Equation 4, increases the coefficient of determination from the previous basic linear regression model that is the CAPM using only the market return as its explanatory variable. The multiple linear regression model includes the explanatory variable avgspread (average spread) and its interaction term with the market return (avgspread x returnSP). The average spread is shown to be statistically significant in the model for Apple Inc. stock and in the robustness tests. The reason that this extra explanatory variable is significant could be due to the fact that the average spread is a measure of the liquidity of the stock. As such, if the average spread variable is lower (meaning the difference between the highest ask price and lowest bid price is subsequently smaller) this means that there is more liquidity in the stock and increased liquidity would affect the return on the stock.

Furthermore, the interaction term (avgspread x returnSP) in the multiple linear regression model is the second most statistically significant variable in the model (after the S&P returns variable) in Table 6 Panel C. This could be due to the fact that if the market (S&P 500) is doing well then the average spread variable is lower (meaning the difference between the highest ask price and lowest bid price is subsequently smaller) because there is more liquidity in the stock and increased liquidity could be a result of the market performing well. This in turn, would affect the return on the stock; hence, the relationship between the average spread and the returns on the stock is affected by the returns on the market.

Accordingly, it can be concluded from this study that a measure of liquidity and its interaction with the market returns are statistically significant explanatory variables in a multiple linear regression model of the CAPM. Therefore, the multiple linear regression model produced in Equation 4 was shown to increase the multiple R^2 value from the basic linear regression model and the extra explanatory variables were shown to be statistically significant in both Table 6 Panel C and Table 7. As such, the multiple linear regression model of the CAPM produced in this paper can increase the amount of variability in the returns on an asset that can be explained as compared to the basic linear regression model of the CAPM. This is an important modification that can better help forecast returns on an asset.

Further research can be conducted using other S&P 500 stocks (other than Apple Inc.) and potentially other stocks from other markets, with their respective market returns to examine if this model (Equation 4) still holds. Additionally, future research can conduct more sectional testing by using less data points. For example, taking at random, a certain amount of days each year rather than taking all daily values or the first twenty daily values of the year as was conducted for Apple Inc. stock in the robustness testing section. Looking at other time periods to test this model is also another route for future research. One of the limitations of this study is that it used daily returns rather than monthly returns due to the spread values being daily values rather than monthly values.

References

CRSP. (2017a). Center for Research in Securities Prices. Available online at http://www.crsp.com/products/documentation/data-definitions-v

CRSP. (2017b). Center for Research in Securities Prices. Available online at http://www.crsp.com/products/documentation/data-definitions-s#shares outstanding

CRSP. (2017c). Center for Research in Securities Prices. Available online at http://www.crsp.com/products/documentation/data-definitions-s#spread-between-big-and-ask

Fama, E. & French, K. (2004). The Capital Asset Pricing Model: Theory and Evidence. *Journal of Economic Perspectives, 18*(3), 25-46. https://doi.org/10.1257/0895330042162430

Frees, E. (2010). *Regression modeling with actuarial and financial applications.* New York: Cambridge University Press, 150-151.

Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *Review of Economics and Statistics, 47*(1), 13–37. https://doi.org/10.2307/1924119

Markowitz, H. (1952). Portfolio Selection. *Journal of Finance, 7*(1), 77-91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x

Markowitz, H. (1959). Portfolio selection: efficient diversification of investments. Cowles Foundation Monograph No. 16. New York: John Wiley & Sons, Inc.

Sharpe, W. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance, 19*(3), 425–442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x