Chapter 3
"An Analytical Approach to
Investments, Finance and Credit"

RISK, RETURN AND MARKET COMPARATIVE ANALYSIS

Beta Coefficient, R-Squared, and Regression Analysis

Regression Analysis: An Analytical Introduction

- **Regression analysis is one of the most fundamental methods used by statisticians to compare the movement of one or many variables to each other**.
- This is different than correlation that also calculates periodic change of one value to another.
- Where the correlation measured from negative -1 to positive 1 designating the direction of the one value to another the beta coefficient is different.
- Beta measures the sensitivity of the movement of the value or the portfolio of stocks to another value such as the benchmark or a market index

Introduction to Linear Regression and Beta Coefficient

The objective for using a linear regression model in statistics is to analyze the straight-line relationship between two variables. The relationship can be explained as follows:

$$y = \beta_0 + \beta_1 x + e$$

where y is the dependent variable or response variable, x is the independent variable or explanatory variable, and β_0.*is* the y-intercept. The y-intercept is the value of y when x = 0, and β_1 *is* the slope of the line. β_1 gives the amount of change in y for every unit change in x. Finally, e is the random error. The random error is included under the premise that analysts are never perfectly precise in their analyses of predicting future outcomes. Many financial models that use regression analysis to predict future values or returns do not include the error calculation. The regression analysis on Excel uses the statistical concept of least squares, a method that seeks the coefficients β_0 and β_1 such as that

$$y = \beta_0 + \beta_1 x$$

Beta Coefficient

Figure 3.1 demonstrates the relationship of various portfolios of stocks as compared to the market index. At beta 1, the portfolio moves exactly at the same pace as the market. Portfolios of stocks that have beta of 2 are moving up and down twice as much as the market.

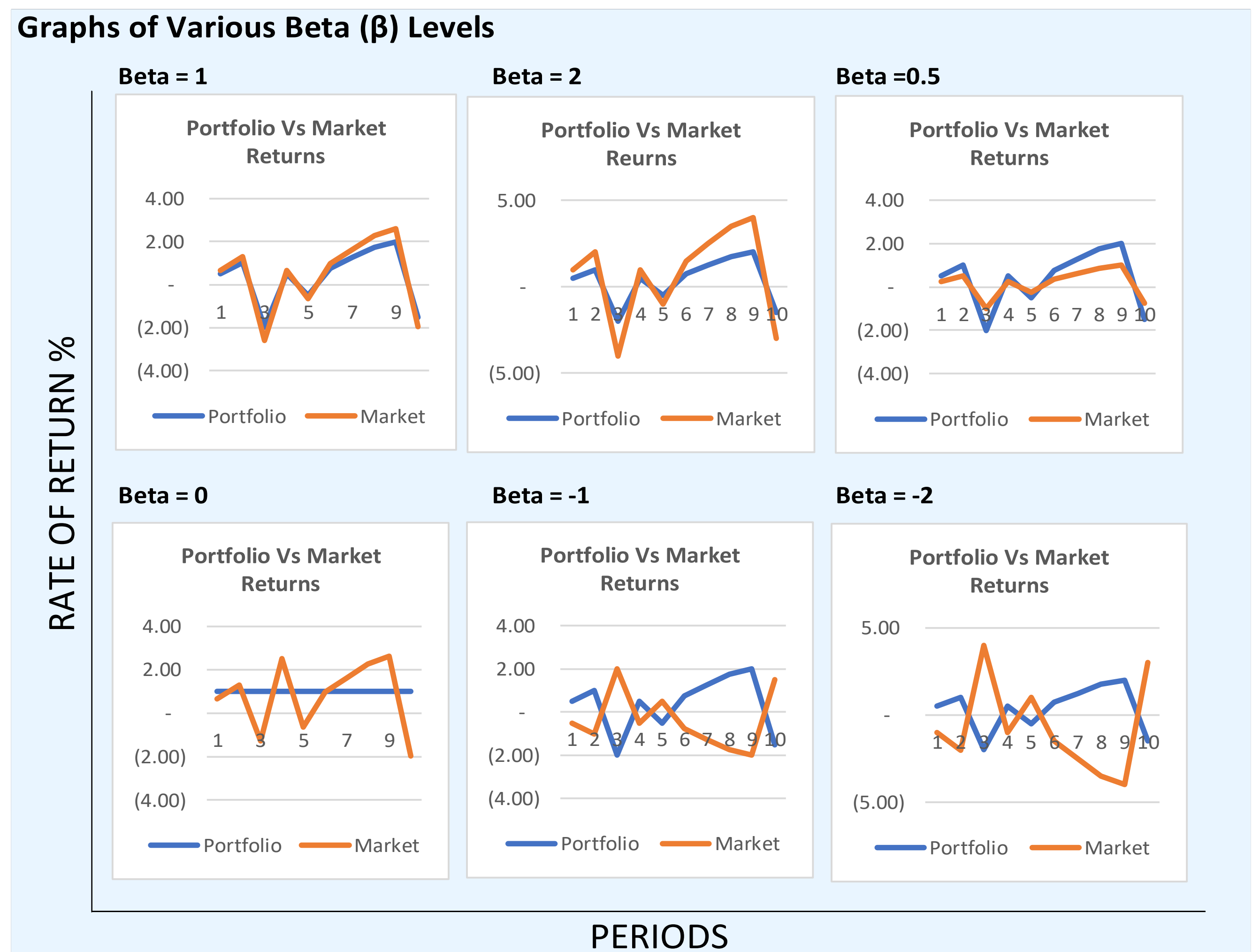

Figure 3.1

Beta Coefficient

- An example of a historical performance of a stock as compared to the performance of the stock market, using the S&P 500 Index, is demonstrated in Figure 3.2.

HISTORICAL ANALYSIS

	Returns		Deviations from Average Return		Product from Deviation	Product from Return
	y	x	$y-\bar{y}$	$x-\bar{x}$	$(y-\bar{y})\cdot(x-\bar{x})$	$(x-\bar{x})^2$
	Stocks %	S&P Index %	Stocks %	S&P Index %	%	%
Year -12	-6.50	-4.60	-18.18	-9.82	178.42	96.37
Year -11	-13.20	-11.30	-24.88	-16.52	410.85	272.80
Year -10	-8.90	-5.00	-20.58	-10.22	210.21	104.38
Year -9	25.00	12.00	13.33	6.78	90.39	46.01
Year -8	48.50	23.00	36.83	17.78	654.87	316.25
Year -7	37.60	25.00	25.93	19.78	512.88	391.38
Year -6	10.50	1.00	-1.18	-4.22	4.95	17.78
Year -5	7.20	4.50	-4.48	-0.72	3.21	0.51
Year -4	-5.60	1.00	-17.28	-4.22	72.84	17.78
Year -3	17.50	3.00	5.83	-2.22	-12.91	4.91
Year -2	21.50	12.00	9.83	6.78	66.65	46.01
Year -1	6.50	2.00	-5.18	-3.22	16.65	10.35
					2209.01	1324.54
Average Return	**11.68**	**5.22**				
Standard Deviation	19.20	10.97			Beta (β)	**1.6678**

$$\frac{\sum[(y-\bar{y})\cdot(x-\bar{x})]}{\sum(x-\bar{x})^2}$$

<u>EXCEL FORMULAS</u>

Slope (β)=	*1.6678*	*Relationship between Dependent Y with Indepent X*
Forecast =	*4.6426*	*Predicts value of y given a value of x=1%*
Standard Error =	*6.0923*	*Predicts the standard error y-value for each x in the regression*

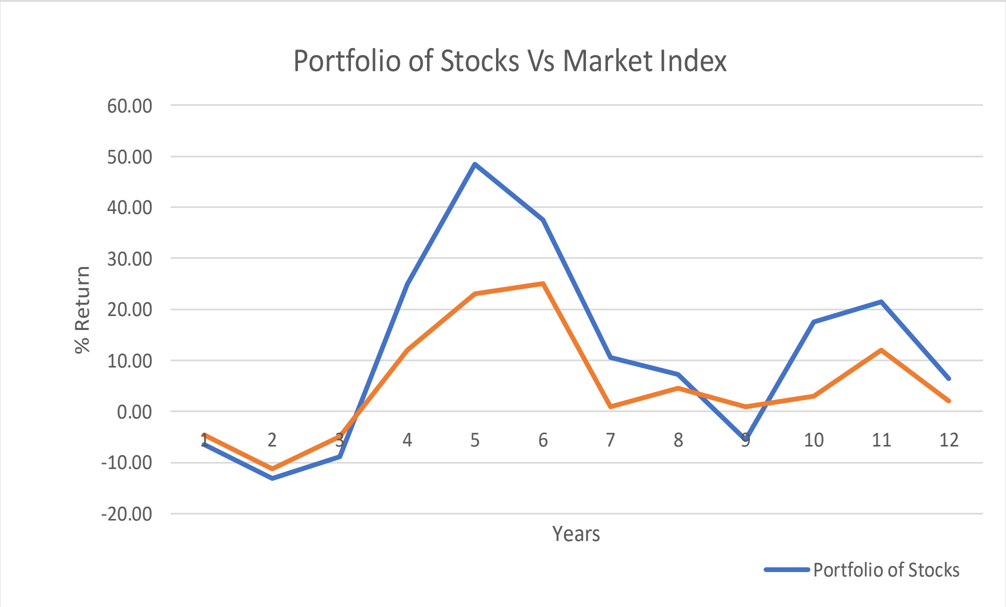

Figure 3.2

Regression and ANOVA

REGRESSION SUMMARY OUTPUT (EXCEL)

Regression Statistics	
Multiple R	0.953139
R Square	0.908474
Adjusted R Square	0.899322
Standard Error	6.092286
Observations	12

Analysis of Variance (ANOVA)					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	3684.083055	3684.083055	99.25877166	1.64467E-06
Residual	10	371.1594445	37.11594445		
Total	11	4055.2425			

Slope (beta)	***Coefficients***	***Standard Error***	***t-Stat***	***P-value***	***Lower 95%***	***Upper 95%***	***Lower 95.0%***	***Upper 95.0%***
Intercept	2.974868822	1.963560699	1.515037871	0.160712968	-1.40021706	7.349954704	-1.40021706	7.349954704
X Variable 1 (Beta Slope)	1.667756775	0.167397229	9.96286965	1.64467E-06	1.294772505	2.040741046	1.294772505	2.040741046

Figure 3.3

Regression Analysis, R-squared, and the Analysis of Variance (ANOVA) Using Excel

The Excel regression analysis yields three-part outputs:

- **Regression statistics:** R-squared and the standard error.
- **Analysis of variance (ANOVA) between the two variables:** The ANOVA shows the degrees of freedom (df), calculates the sum of squares (SS), the mean squares, the F statistic, and significance F. The ANOVA shows whether there are any statistically significant differences between the means of two or more variables.
- **Regression coefficients:** The slope or beta coefficient, the t test and upper/lower 95% limits “within” the group variability and “between” the group variability. The t test is another measurement for comparing the samples. The t test is more effective if there is a comparison of two variables.

Regression Statistics

- ***Multiple R:*** **The multiple R, which is measured between 0 to 1, reflects the correlation coefficient of the two variables. It tells the analyst how strong the linear regression**
- ***R-squared*:** R-squared (r^2), also called coefficient of determination, measures how many points fall on the regression line between the two variables.
- ***Adjusted R-squared:*** This number reflects a more accurate r^2 if the number of variables used are more than one.
- ***Standard error:*** The standard error in a regression analysis measures how precise the regression coefficient is. In other words, it tells the analyst how spread out the y variables are around the mean (average).
- ***Observations:*** Number of observations in the analysis. Figure 3.3 shows 12 representing the 12 years of historical performance for both the portfolio and market index annual returns

ANOVA

- ***Degrees of freedom (df):*** When calculating historical data, the average standard deviation of each of the variables is based on the number of observations minus one (n - 1).
- ***Between-group variability and sum of squares:*** When looking for comparing two samples—in this case, the portfolio value and market index—there could be an overlap in the date.
- ***Within-group variability and mean squares:*** If the analyst plots the data into a normal distribution and compares each variable to the grand mean within the two variables it could calculate the within-group variability, which is the distance between the two means.
- ***F statistic or significance F:*** The F statistic or significance F and sometimes called the F ratio, is the number that measures if the means of different samples are significantly different. The lower the number, the more similar the sample means are.

ANOVA

ANALYSIS OF VARIABLES (ANOVA)

Returns		Sorted	
y	x	y	x
Stocks %	S&P Index %	Stocks %	S&P Index %
-6.50	-4.60	-13.20	-11.30
-13.20	-11.30	-8.90	-5.00
-8.90	-5.00	-6.50	-4.60
25.00	12.00	-5.60	1.00
48.50	23.00	6.50	1.00
37.60	25.00	7.20	2.00
10.50	1.00	10.50	3.00
7.20	4.50	17.50	4.50
-5.60	1.00	21.50	12.00
17.50	3.00	25.00	12.00
21.50	12.00	37.60	23.00
6.50	2.00	48.50	25.00

11.68	**5.22**	**Average Return**
19.20	10.97	Standard Deviation

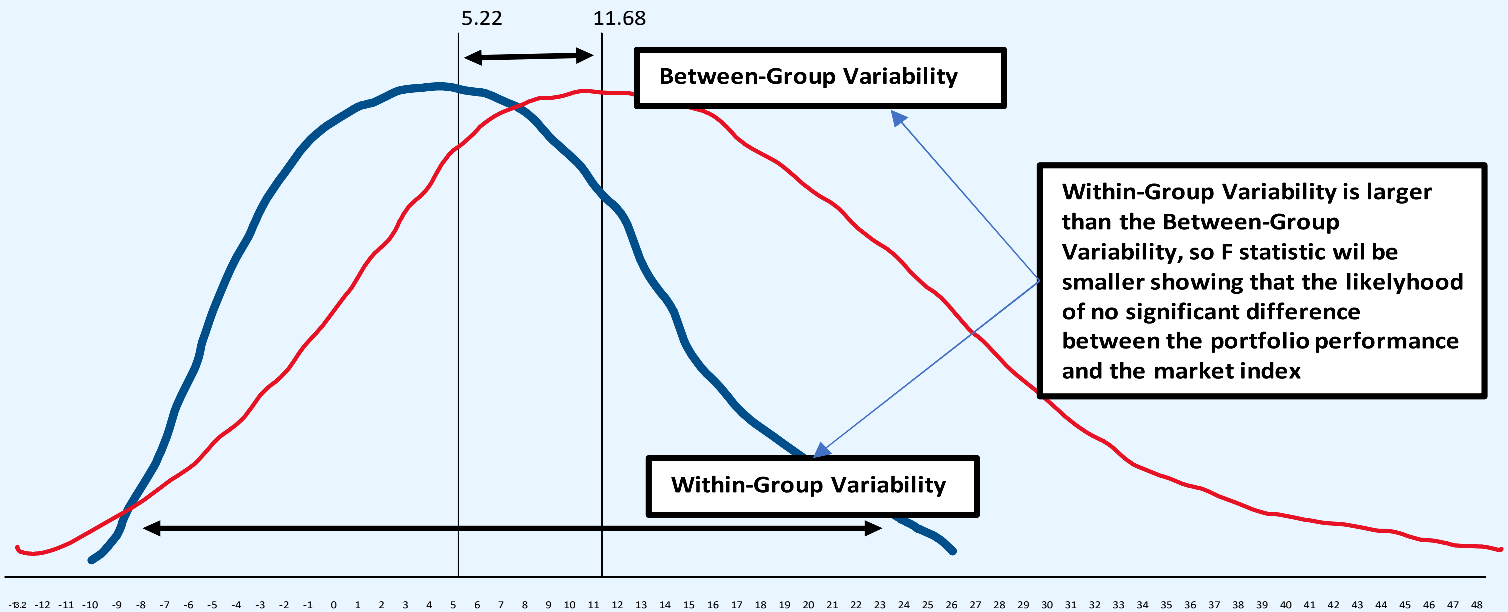

Figure 3.4

Regression Coefficients

- ***Beta coefficient (slope):*** As mentioned, the linear regression analysis is basically the linear relationship between the independent (x) and dependent (y) variables expressed in the equation $y = \alpha + \beta x$, called the regression line. The first calculation coefficient is beta coefficient (β), also known as the slope.
- ***Standard error:*** The standard error is similar to standard deviation measuring the spread or the deviation from the data's average. Though both are calculated the same, the standard error uses smaller statistical sample data, and the standard deviation uses the entire population data (aka parameters) to calculate the difference from the average.
- ***t test:*** The t test or t statistic is used when the analyst decides to accept or reject the null hypothesis—basically that the movements or performance of the dependent variable (y) do depend on the movement of the independent variable (x).
- ***p value:*** The p value, or probability value, is expressed from 0-1 and it measures the percentage difference in the combined average of the variables.
- ***Lower and upper 95% confidence intervals:*** The confidence intervals are expressed as a percentage, in this case 95%. It means that the results are expected to match 95% of the time.

Regression Coefficients

- ***Beta coefficient (slope):*** As mentioned, the linear regression analysis is basically the linear relationship between the independent (x) and dependent (y) variables expressed in the equation $y = \alpha + \beta x$, called the regression line. The first calculation coefficient is beta coefficient (β), also known as the slope.
- ***Standard error:*** The standard error is similar to standard deviation measuring the spread or the deviation from the data's average. Though both are calculated the same, the standard error uses smaller statistical sample data, and the standard deviation uses the entire population data (aka parameters) to calculate the difference from the average.
- ***t test:*** The t test or t statistic is used when the analyst decides to accept or reject the null hypothesis—basically that the movements or performance of the dependent variable (y) do depend on the movement of the independent variable (x).
- ***p value:*** The p value, or probability value, is expressed from 0-1 and it measures the percentage difference in the combined average of the variables.
- ***Lower and upper 95% confidence intervals:*** The confidence intervals are expressed as a percentage, in this case 95%. It means that the results are expected to match 95% of the time.

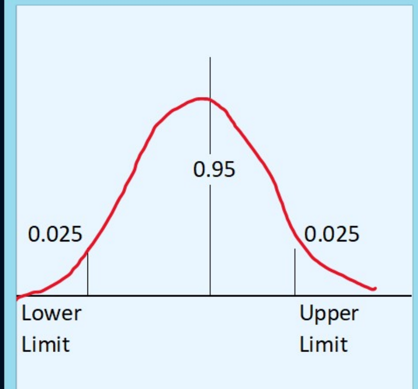

Figure 3.5

Excel Formulas

Excel Formulas Reviewed:

- Annual rate of return: = IRR (Range of cash flows including the initial investment as negative)
- Average of returns: = AVERAGE (Range of periodic returns)
- Variance: VAR (Range of data)
- Standard deviation: STDEV (Range of data)
- Covariance: COVARIANCE.P (Range 1, Range 2)
- Correlation: = CORREL (Range 1, Range 2)
- R-squared: = RSQ (Range 1, Range 2)

Analysis of Variance (ANOVA) and Regression Analysis Using Excel:

- Data tab / Analysis group, click Data Analysis / Regression Analysis, then select the INPUT Y Range (Dependent Variables) and INPUT X Range (Independent Variable)