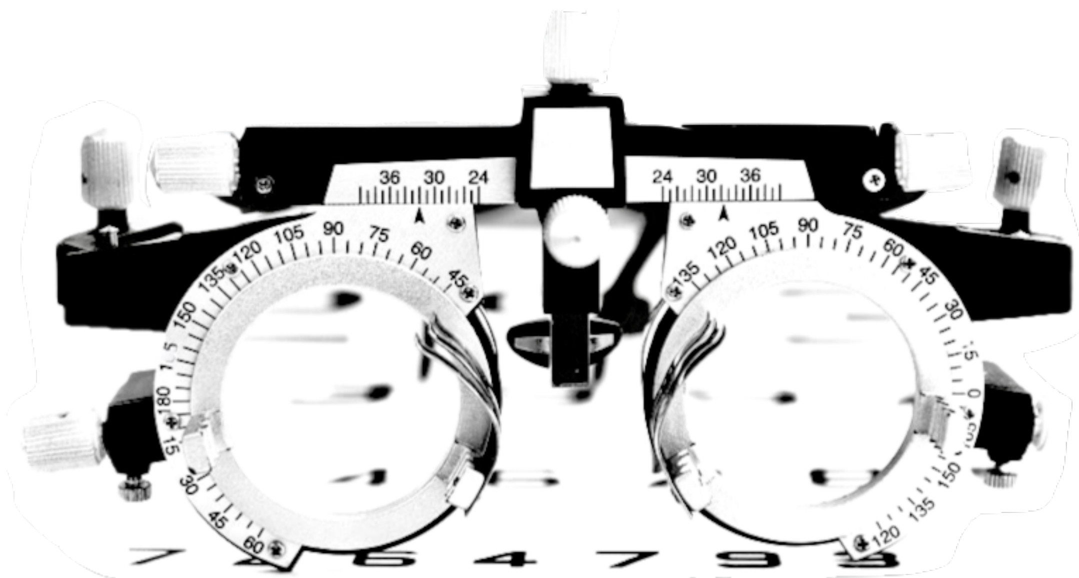
NITH INVESTMENTS
TECHNOLOGY • INSIGHT • CLARITY

MULTIPLE LINEAR REGRESSION

A Closer Look at Our Equity Models



Sean Deutsch, CFA
Director, Investment Research

TABLE OF CONTENTS

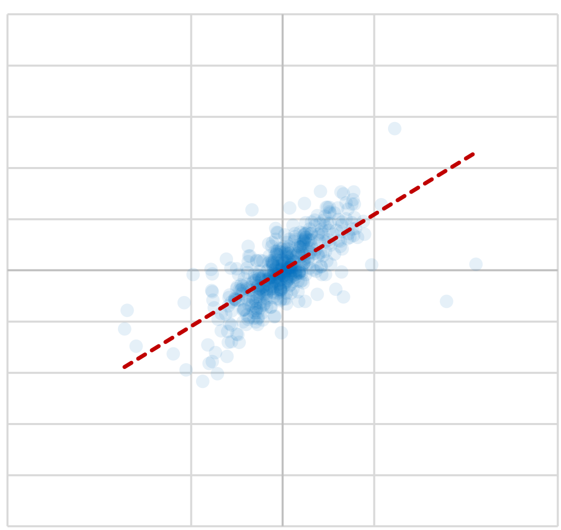
INTRODUCTION

The equity pricing models at Nith Investments are driven by a linear regression approach to forecast expected returns of individual stocks. With over 20 years of investment analytics experience, we have found that linear regression, specifically multiple linear regression, is the most practical and transparent model for understanding stock returns. Most importantly, it is the most easily explainable model for most of our clients' use cases.

For interested subscribers to our mailing list, we can send a complimentary paper on how we combine our regression models with Monte Carlo simulation to forecast a portfolio of stocks.

MULTIPLE LINEAR REGRESSION

Regression models are used to describe the relationship between variables by fitting a line to observable data. In a simple linear regression, where one independent variable (x) is used to predict another variable (y), we can plot all the (x, y) pairs that have already been observed in history on a Cartesian graph.



We then fit a line through that data such that the deviation of each observation to that line is minimized.

The line of best fit (the red line above) can be summarized by this formula:

Simple linear regression formula

$y_i = \beta_0 + \beta x_i + \epsilon$

where, for $i = n$ observations:

y_i = dependent variable
x_i = explanatory variables
β_0 = y-intercept (constant term)
βx_i = slope coefficients for the independent variable
ϵ = the model's error term (also known as the residuals)

For most regression analysis based on stock returns, we focus on βx_i -- the slope coefficient, also known as beta. For example, if the beta is 1.3, then we can reasonably expect the dependent variable to move around +1.3% when the independent variable moves +1.0%.

The pricing function used in our equity models is based on a multiple linear regression, also known simply as multiple regression. Multiple regression is an extension of simple linear regression discussed above. This statistical technique models the linear relationship between several independent variables (sectors, factors, etc.) and the dependent variable (a stock). Extending to multiple regression allows the model to take inputs from several independent variables so it can have better explanatory power.

The extension from a simple linear regression to a multiple linear regression can be summarized by this formula:

Multiple linear regression formula

$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \epsilon$

where, for $i = n$ observations:

y_i = dependent variable
x_i = explanatory variables
β_0 = y-intercept (constant term)
β_p = slope coefficients for each indpendent variable
ϵ = the model's error term (also known as the residuals)

EQUITY MODELS

The equity pricing models at Nith Investments perform a multiple regression based on a stock's country of risk and the primary sectors and indices of that country. The primary sectors and indices are used as the independent variables, and at least two years of daily returns are required for the regression to calculate a meaningful output. For IPOs, clients can proxy the stock to a major market index, ETF, or an individual equity with a similar risk profile.

*Primary sectors used in the multiple regression for stocks with the USA as their country of risk

The betas calculated from the regression can also be used to show the P&L during historical crisis events like Covid-19 or the financial crisis of 2008. The advantage of using the betas from independent variables is that stocks do not need pricing data during the actual stressed time-period. For example, using the betas calculated from the model we can see how Google would perform during 9/11 or the Russian Debt Crisis even though it only started trading in 2004.

EXAMPLE: AMERICAN EXPRESS (AXP)

Using two years of daily closing prices for American Express (AXP) as of October 31, 2023, and the corresponding daily index levels, we can perform a multiple regression for AXP and identify how it would react to changes in the broader market. The regression is performed on the daily returns calculated from the daily AXP prices and index levels.

Fitting a hypothetical line though the data, the regression model returns the following output:

Adjusted R-Squared: 0.644

Significance F: 0.000

	Coefficients (beta)	P-value
S&P 500	-0.013	0.955
Russell 2000	0.062	0.496
US Basic Materials	-0.004	0.963
US Consumer Goods	0.100	0.201
US Financial	1.334	0.000
US Health Care	-0.311	0.001
US Industrials	-0.010	0.939
US Energy	-0.002	0.952
US Technology	0.010	0.905
US Telecommunications	-0.079	0.105
Intercept	0.000	

The first output, the Adjusted R-squared, measures how much of AXP's variation is explained by the sectors and indices in the regression. Here, we see that over 64% of the variation is explained by the independent variables. In the social sciences, an R-squared above 0.6 roughly implies a high level of correlation and an R-squared below 0.4 implies a low level of correlation.

The Significance F tells us whether the overall regression is statistically significant, where a lower value implies higher significance. In the social sciences, most practitioners regard values below 0.05 as significant.

For our purposes, we rely on the beta outputs to predict stock returns. In the table above, we can interpret the beta of each sector as the sensitivity of AXP to that sector.

Recall the multiple regression formula, where a stock's return is the sum of its sensitivity to each independent variable:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \epsilon$$

Using the regression results, this translates to:

Return of AXP =

S&P 500 x -0.013 +

Russell 2000 x 0.062 +

US Basic Materials x -0.004 +

US Consumer Goods x 0.100 +

US Financial x 1.334 +

US Health Care x -0.311 +

US Industrials x -0.010 +

US Energy x -0.002 +

US Technology x 0.010 +

US Telecommunications x -0.079,

where the bold text refers to the return of each index.

As expected, we see that AXP is more sensitive to US Financial than to any other sector with a beta of 1.334 to that sector (highlighted in green).

The P-value of each sector indicates how significant that sector is in explaining the sensitivity. Like the Significant F statistic, a lower value indicates a higher significance.

PREDICTING A PORTFOLIO OF STOCKS AND STRESS TESTING

Equipped with the regression equation, we can predict the return of AXP stock by simulating the return of each sector. Since the sectors and major indices are common market factors, their simulated returns can be applied simultaneously to all other stocks that have the USA as their country of risk. In a separate paper, we explain how Monte Carlo simulation can be used to simulate the returns of the common factors and applied to a portfolio of stocks.

Additionally, we can conduct what-if analysis using the betas from the regression output. A good example of a what-if analysis is historical stress testing. Since we already know how each sector and index performed during a historically stressed event, we can apply those returns to the regression equation and compute a hypothetical response a stock would make if the stressed event occurred again.

CONCLUSION

From our experience, multiple linear regression has been very useful in explaining stock performance. Hopefully, it is apparent that this model intuitively deconstructs stock returns and can be used to perform sensitivity analysis such as historical stress testing.

If you would like to obtain more details or provide your comments on this paper, our contact information can be found here: www.nithinvestments.com/contact