A Comparison of Linear and Nonlinear Equity Factor Models: Evidence from the UK Market

April 2020

University of Bath
Department of Economics

Acknowledgements

I extend my thanks to Andrzej for providing me with his expertise on equity factors, Patrick for his knowledge on statistics and machine learning, and Dad for his help in structuring my writing. Your advice and comments are all greatly appreciated.

Abstract

This thesis investigates the extent to which the addition of nonlinearities to traditional equity factor models can improve upon the performance of a linear OLS specification in both fitting the return relationship and forecasting future security returns. These nonlinearities are introduced via cubic predictor sets and supervised machine learning models. While the inclusion of cubic powers is found to add more in complexity than performance, the machine learning models provide some additional modelling and predictive capability. Ridge regressions are found to improve upon the standard OLS model in fitting a relationship between common factor models and security returns. The reason for this is suggested to be the nonlinearity of security returns in relation to equity risk premia, where returns in the tails of the distribution experience a different relationship to those in the centre, distorting the standard linear model. Finally, regression trees are found to be the most performant in solving the prediction problem, suggesting value of nonparametric models in forecasting security returns. Despite this, the results are unlikely to have economic significance in terms of profitability after trading costs are taken into account.

Table of Contents

1. Introduction

For as long as the stock market has existed, its participants have been searching for ways to accurately assess the expected price return of a stock. The main motivation is usually to make money, but beyond this a security's expected return can be used by analysts to estimate the cost of capital of a company and hence determine its discounted fundamental value today.

The theory of equity risk premia is that the expected return of a stock can be decomposed into a contemporaneous function of various equity factor exposures. The generalised linear specification is shown in equation 1.

$$E(r_{i,t}) = \alpha_{i,t} + \sum_{j=1}^{n} \beta_{i,j} F_{j,t} + u_{i,t} \tag{1}$$

Where each $F_{j,t}$ is the return of factor j in period t, and $\beta_{i,j}$ is the exposure of stock i to factor j. For a factor model to be complete, $\alpha_{i,t}$ must be statistically indifferent from 0, and $u_{i,t}$ must be independent and identically distributed (i.i.d) with a normal distribution. Together these two facts imply that there is no autocorrelation in the error terms, hence no information could be added to improve the model, and that the model captures the entirety of an asset's return. As a result, investors can tilt their portfolios toward companies with higher factor exposures, increasing the expected return of their portfolio as a result. Equity factors are generally modelled in this contemporaneous manner, with the goal of explaining the return of a security via the return of a set of common factors. This thesis looks at the problem in the same way, subsequently extending it to the widely known problem of return forecasting as a corollary.

There are three broad causes of return factors (Pioch, 2018). Firstly, there is compensation for risk, in which an investor earns a higher expected return in exchange for taking more risk. For example, the illiquidity risk of owning a small company leading to the size effect. Secondly, there are factors caused by behavioural biases of market participants. Our tendency to overweight recent observations, disproportionately fear losses and idolise a compelling story lead to some of the persistent inefficiencies that we observe in the markets today. Thirdly, equity risk premia can be a product of market structure. An example of this is the January effect, where fund managers sell stock in December to lock in their gains, then buy it back in January in order to make returns for the new year, leading to lower returns in December and higher returns in January.

As will be elaborated upon in the following section, equity factors have generally been modelled and tested using a linear specification of Ordinary Least Squares (OLS) such as the one shown above. The simplicity of the linear model makes it easy to interpret and hence an attractive choice for the estimation of equity factors, but that simplicity may in turn sacrifice accuracy. Consequently, a model that allows for nonlinearity can potentially offer a better fit to the true relationship better. This can be done in one of three ways: Firstly, by taking a linear model and providing it with a nonlinear predictor set, secondly, by passing a linear set of predictors into a nonlinear model, or finally by passing a nonlinear set of variables into a nonlinear model. These three combinations can be compared to the linear model using a linear predictor set to see if the accommodation of nonlinearities yields any benefit to the researcher, statistician or investor.

Some analysis of this type has been done over the years, with the idea for this thesis being borne out of a recent publication by Gu, Kelly and Xiu (2019). Given the larger dataset in the US, most of the research on this topic has been conducted there. A corollary goal of this thesis is to provide the body of research with an attempt to do the same using UK data, allowing for a comparison of nonlinear multi factor models across the two markets. It should be noted that UK data has a shorter, more recent time scale and less detail, which this thesis acknowledges.

The rest of this paper is broken down into 5 main sections. Section 2 provides a summary of the research that has been done on equity factors and their nonlinearity. Section 3 elaborates on the data upon which the research was conducted, as well as the main calculations made. Section 4 details the models fitted, including how they are optimised and evaluated. Section 5 concludes with the key findings when applying these models to the dataset in question. Finally, section 6 extends the idea from contemporaneous regressions, outlining the performance of each model in forward looking prediction.

2. Literature Review & Theory

The literature for this thesis originates from two main fields, namely, equity risk premia and machine learning. After a brief summary of the efficient market hypothesis, section 2.2 outlines the main models that have been researched in the field of equity factors, while section 2.3 summarises the literature on machine learning and its application to modelling risk premia.

2.1. Efficient Market Hypothesis

As a precursor to the literature on equity factors, it is first important to acknowledge the implications for market efficiency. Eugene Fama (1970) proposed the idea of the efficient market hypothesis (EMH) in which he posited that at any given moment, prices will fully reflect all information, private and public, that is available to the market. Consequently, assuming that the discovery of new information is a random process, stock returns must also follow a random process.

Depending on the driver of excess returns, an equity factor may or may not contradict the EMH. Of the three drivers of risk premia mentioned above, compensation for risk is undoubtedly in line with the EMH given the risk-return trade-off of the markets. The other two drivers however create the potential for market participants to achieve risk adjusted returns that are consistently higher than the market, and therefore contradict the EMH.

The implications of the EMH for equity factors are that no investor, statistician or trader can use their set of available information to correctly and consistently forecast asset returns to beat the market.

2.2. Equity Risk Premia

2.2.1. The Market

The first model that proposes stock returns are composed of a set of factors is the Capital Asset Pricing Model (CAPM) put forward independently by William Sharpe (1964) and John Lintner (1965). It states that the excess return of a stock above that of the risk-free rate is equal to a stock specific measure of risk, beta, multiplied by the risk premium of the stock above that of the market.

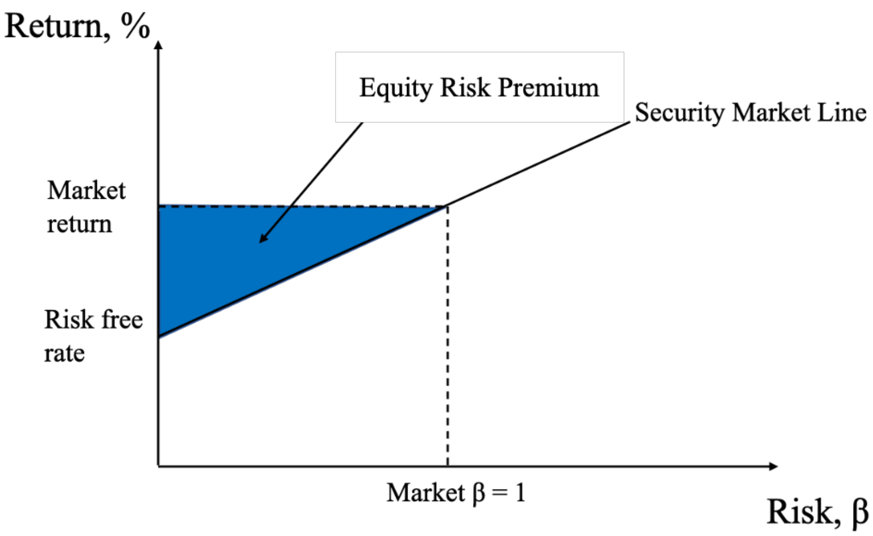
$$r_{i,t} - r_{f,t} = \beta\left(r_{i,t} - r_{m,t}\right) \tag{2}$$

The beta of a stock represents its riskiness relative to the market and can be defined by equation 3 below, or more commonly by the regression coefficient produced when running OLS on equation 2.

$$\beta = \frac{cov(r_i, r_m)}{var(r_i)} \tag{3}$$

In the risk-reward space, the CAPM model is represented by the Security Market Line (SML), depicted in figure 1. The upward slope of this line indicates that by taking on additional risk an investor can reasonably expect to earn a higher return. By taking the risk of the market as defined by a beta of 1, figure 1 graphically represents what is known as the equity risk premium, which equals the expected return on the stock market in excess of the risk-free rate.

Figure 1: The Equity Risk Premium



The existence of the such a premium is in accordance with the EMH, because investors are rewarded with a higher expected return for taking the additional risk of putting their money into the stock market, which is the highest risk asset class.

2.2.2. Size and Value

A few decades after Sharpe and Lintner, Eugene Fama and Kenneth French (1993) introduced the concept of multiple equity factors by suggesting that stock returns are driven by more than just the equity risk premium. In their three-factor model, Fama and French (hereby FF) showed that as well as the market risk premium, stocks are also exposed to the size factor and value factor.

FF define *the Size factor* as the difference in average return on the smallest and largest 30% of stocks, resulting in the "Small Minus Big" (SMB) factor. There are two common explanations for the outperformance of smaller companies (Size factor definition - Risk.net, 2020). Firstly, smaller companies are less liquid, and so in buying shares in the company, an investor is accepting the risk that they may not be able to sell them for a fair price in the future if demand dries up. Secondly, smaller companies are viewed as inherently riskier as they are more susceptible to changes in the economy and business cycle. Either way, the size factor would indicate investors being compensated for taking additional risk.

T*he Value factor* is calculated by FF as the difference in average return between companies with the highest and lowest book-to-market value, giving the "High Minus Low" (HML) factor. The idea of value investing is that cheap stocks tend to outperform expensive ones in the long run. There are various definitions of value in a company, including dividend yield, book to market value and price to book value. However, all indicate the same characteristic, namely, that the investor is getting a lot for what they are paying.

FF found that, in their dataset, over 90% of the variation in security returns could be explained by the three factors in question. The FF (1993) three-factor model is specified in equation 4.

$$r_{i,t} - r_{f,t} = \beta_1 MKT_t + \beta_2 SMB_t + \beta_3 HML_t + u_{i,t} \tag{4}$$

2.2.3. Momentum

Momentum is the observation that stocks with strong recent past performance will continue to outperform in the future, meanwhile those that performed poorly will continue to underperform. The idea that equity returns possess this serial autocorrelation was first put forward by Jegadeesh (1990), who found a negative autocorrelation lags as short as one month, but positive correlation at longer lags, especially twelve months.

The prevailing reason behind the existence of momentum is that humans participate in the stock market, and humans have behavioural biases (Carlson, 2020). There are specifically two biases which are broadly accepted as contributing to the momentum factor. Firstly, there is recency bias. This says that people put too much value on recent observations when making decisions, including investment decisions. Consequently, stocks that have performed well recently will look like good investments and hence perform better in the future as more investors buy into the story. Secondly, individuals often suffer from social proof, or herding (Bikhchandani and Sharma, 2000). Typically, when a person is facing a tough decision, they look to the actions of others to discover the best choice. If the price of a stock has been rising, that means that other people have been buying it, so in order to follow what others are doing an investor should also buy the stock too, driving up the price.

Jegadeesh's 1990 paper was built on by two more influential papers. Firstly, Jegadeesh and Titman (1993) evaluated the consequences of momentum for stock market efficiency. They conclude that returns from momentum trading strategies are not due to a risk-reward relationship and hence are in contrast to the EMH.

Secondly, Mark Carhart (1997) proposed the *Carhart four-factor model* a few years later, which adds a momentum factor to the FF three-factor model by including an Up Minus Down (UMD) factor, which equals the average return of past winners minus the average return of past losers.

$$r_{i,t} - r_{f,t} = \beta_1 MKT_{\mathrm{t}} + \beta_2 SMB_{\mathrm{t}} + \beta_3 HML_{\mathrm{t}} + \beta_4 UMD_{\mathrm{t}} + u_{i,t} \tag{5}$$

The Carhart (1997) four-factor model, as well as the papers of Jegadeesh and Titman (1990, 1993), are highly significant in the literature of equity risk premia, primarily because momentum has been the most prominent and profitable of the commonly researched factors, while also being a significant explanatory variable in the returns of fund managers throughout history (Carhart, 1997).

2.2.4. Quality

Many years later, Fama and French (2015) added to their three-factor model by incorporating two Quality factors. Their assertion is that managers and investors will favour high quality companies who invest well and make solid profits.

Firstly, *low investment* factor, CMA, is the difference between the return of companies that invest conservatively and the return of companies that invest aggressively, as defined by asset growth. The intuition is that companies that do not invest a lot have higher standards for projects that they take part in, ensuring a higher rate of return as they are not experiencing the same level of diminishing marginal returns that those companies investing heavily are.

Secondly, *profitability* is defined by the RMW factor, which takes the return of companies with robust profits minus the returns of companies with weak profits. More profitable companies will intuitively earn a higher return as investors pay more for a share of the higher profits. However, this does not explain why the EMH is not functioning here, and hence why any excess return is not bought up immediately.

The two quality factors can be explained using the Dividend Discount Model (DDM) of asset pricing (Miller and Modigliani, 1961), which posits that the current value of a company is equal to the discounted present value of its expected future dividends. Note here that M_t is the market value of a security, $E(D_t)$ it the expected dividend in period t, r is the discount rate, B_t is the book value of the company and Y_t is the company's earnings.

$$M_t = \sum_{t=1}^{\infty} \frac{E(D_t)}{(1+r)^t} \tag{6}$$

Dividends can be defined as company earnings minus the change in book value of the company, giving equation 7.

$$M_t = \sum_{t=1}^{\infty} \frac{E(Y_t - \Delta B_t)}{(1+r)^t} \tag{7}$$

Finally, in accordance with FF (2015), dividing through by current book value gives equation 8.

$$\frac{M_t}{B_t} = \sum_{\tau=1}^{\infty} \frac{E(Y_{t+\tau} - \Delta B_{t+\tau})}{B_t(1+r)^{\tau}} \tag{8}$$

Robust profitability implies a high value of $Y_{t+\tau}$ in equation 8, while conservative investment implies a low $\Delta B_{t+\tau}$. Both lead to a higher value of M_t in equation 7 and M_t/B_t in equation 8.

Together these additional factors make up the FF five-factor model specified in equation 9.

$$r_{i,t} - r_{f,t} = \beta_1 MKT_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 CMA_t + \beta_5 RMW_t + u_{i,t} \tag{9}$$

2.2.5. Volatility

Finally, Andrew Ang (2006) showed that companies have an exposure to market volatility, and those with a lower sensitivity on average have higher risk-adjusted returns in the future. Specifically, Ang used the VIX index, a measure of market volatility, to show that companies with less exposure to changes in the index had Sharpe ratios which in future periods outperformed those higher exposure stocks, to a statistically significant degree.

The main theory behind the existence of the low volatility factor is what is known as the lottery effect (What is Low Volatility and Why Does It Matter? - Invesco, 2020). This says that in order to achieve a higher expected return, investors go looking for more volatile companies that can offer these higher returns by allowing them to take additional risk. Consequently, less volatilite companies are underbought and hence offer a risk-adjusted return premium to investors as a result of being neglected for higher volatility stocks.

In his paper, Ang proposed multiple models for volatility, two of which built off of the CAPM and FF three-factor models respectively, as specified in equations 10 and 11.

$$r_{i,t} - r_{f,t} = \beta_1 MKT_t + \beta_2 VIX_t + u_{i,t} \tag{10}$$

$$r_{i,t} - r_{f,t} = \beta_1 MKT_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 VIX_t + u_{i,t} \tag{11}$$

2.3. Non-Linearity

So far all of the models put forward use a linear specification of OLS to estimate the relationship between the respective factor and equity returns. This section summarises the models and respective papers which relax the assumption of linearity, in most cases finding promising results. As expected, there is less research of this type to be found in academic journals. The success of this research brings with it the prospect of financial gain, a benefit that few would be willing to share.

2.3.1. Generalised Linear Models

The first extension of the linear model which allows for nonlinearity is a family of models which fall under the generalised linear models. Despite the name, these models are linear only in specification. Generalised linear models involve running OLS on functions of the independent variables as opposed to the variables themselves. As the name suggests, this is a highly generalised setup as it allows for any function to be passed. The generic form of the model is specified in equation 12.

$$Y_i = \alpha + \sum_{j=1}^{k} \left(\beta_j f_j(x_i)\right) + u_i \tag{12}$$

As a result, these models allow for nonlinearities between each x_i and Y_i. A simple example is the polynomial regression. Most papers that apply machine learning to equity risk premia modelling also include a polynomial regression or other form of generalised linear model. Most recently, Gu, Kelly and Xiu (2018) investigate how generalised linear models compare to a linear specification, as well as to other machine learning models, finding that they marginally outperform OLS in both monthly and annual return prediction.

2.3.2. Supervised Machine Learning

Building on generalised linear models is common in machine learning literature, with the most common extension being that of supervised machine learning. This type of machine learning involves fitting a model using a defined set of independent variables, while supplying a dependent variable on which to optimise the performance of predictions. In the case of this thesis the independent variables are the equity factors, while the dependent variable is the security return. This method is in contrast to unsupervised machine learning in which no dependent variable is provided. Such methods are generally used for grouping observations and identifying underlying fundamental characteristics of a set of variables. Unsupervised machine learning does not lend itself as well to the problem in this thesis, as it is mostly used to group observations into observable and unobservable categories. As a result, it will not be discussed further.

Of the papers in finance literature regarding equity risk premia, those that implement machine learning have shown promising results. Gu, Kelly and Xiu (2018) investigate the performance of various machine learning models in forecasting expected returns in the US market. They find that machine learning models have stronger predictive power than linear regression, with neural networks and regression trees being the most performant. The authors identify the advantage of relaxing the nonlinearity assumption, as well as the capacity for a larger predictor set as potential reasons for the strong performance of machine learning models. Further to this, they rationalise the strong performance of non-parametric models such as regression trees as evidence of potentially complex interaction effects being present in the true model.

2.3.3. Deep Learning

Beyond traditional machine learning, research is also being conducted in the area of deep learning, which involves using neural networks with multiple hidden layers for prediction. The first to investigate how artificial neural networks (ANN) can be applied to equity factors was Levin (1996), while more recently Nakagawa, Uchida and Aoshima (2018) and Nakagawa (2019) estimate deep factor models of equity returns.

As this thesis falls under the Department of Economics, little will be said of these models, which lie closer to the study of computer science. The reader is encouraged to survey the aforementioned literature for an idea of the state of research in this area. Furthermore, Gu, Kelly and Xiu (2018) find some evidence that deep learning models do not outperform shallow learning or traditional machine learning on their dataset.

3. Data

This section outlines the data used in this thesis. Section 3.1 outlines the sources of the data, as well as how it was aggregated and how certain parts filtered out. Section 3.2 then describes how the factor returns are constructed to replicate those from the original papers using the data from section 3.3. Following this, 3.4 outlines the predictor sets that are used for each model. Finally, section uses the newly constructed factors to provide a rationale for modelling equity factors in a nonlinear fashion.

3.1. Collection & Sources

The data used in this thesis is collected from Thomson Reuters DataStream over the period 31/12/1995 to 31/12/2018. It consists of monthly, stock level data for all companies in the FSTE All-Share, conditional on sufficient data availability. Investment trusts, unit trusts and other investment vehicles are removed due to their anomalous characteristics, which are both undesirable and unintuitive when estimating equity risk premia. All predictors are either downloaded directly from DataStream or calculated using data that is. The full list of downloaded and calculated predictors and time series can be found in appendix A.

Companies are deemed eligible for inclusion in regressions in period t if the independent variables used in the regression are available in period $t-1$, and return on the security (the dependent variable) available in period t. This ensures that for cross-sectional regressions, a company cannot contribute to the matrix of independent variables without also contributing to the vector of the dependent variables. Both security and market returns are calculated in excess of the risk-free rate, which is taken to be the interest rate paid on 3-month UK government bonds ("Gilts"). This means that regressions implicitly control for the rate of interest in the UK at the time.

Finally, it is worth noting that the dataset used in this thesis is highly dissimilar to data used in common research, in that it is data from the UK market, on a shorter and much more modern time horizon. As a result, this thesis will also be able to evaluate the extent to which traditional models of equity factors hold in situations alternate to the US before the turn of the millennium.

3.2. Constructing Factor Returns

Before testing various models, factor returns must be constructed in a similar fashion as the papers from which they originate. The factors used in this thesis encompass all those included in the section 2.2. Table 1 details how each factor is calculated. Note that while FF tend to take the top and bottom 30% of companies, here only the top and bottom quintile is used to allow for more concentrated factor returns.

Table 1: Definitions of Factor Returns

Market	$= Excess\ return\ of\ the\ market\ (FTSE\ AllShare)$
Value	$= Return\ of\ 20\%\ of\ companies\ with\ highest\ book\ to\ market\ value$ $- Return\ of\ 20\%\ of\ companies\ with\ lowest\ book\ to\ market\ value.$
Size	$= Return\ of\ 20\%\ of\ companies\ with\ smallest\ market\ value$ $- Return\ of\ 20\%\ of\ companies\ with\ highest\ market\ value.$
Low Investment	$= Return\ of\ 20\%\ of\ companies\ with\ highest\ return\ on\ invested\ capital$ $- Return\ of\ 20\%\ of\ companies\ with\ lowest\ return\ on\ invested\ capital$

Profitability	$= Return\ of\ 20\%\ of\ companies\ with\ highest\ operating\ profit\ margin$ $- Return\ of\ 20\%\ of\ companies\ with\ lowest\ operating\ profit\ margin$
Momentum	$= Return\ of\ 20\%\ of\ companies\ with\ highest\ last\ 12\ month\ price\ return$ $- Return\ of\ 20\%\ of\ companies\ with\ lowest\ last\ 12\ month\ price\ return$
Volatility	$= Return\ of\ 20\%\ of\ companies\ with\ lowest\ 24\ month\ return\ volatility$ $- Return\ of\ 20\%\ of\ companies\ with\ highest\ 24\ month\ return\ volatility$

Given that not all data used in the original papers is available in the UK, in some cases the method of calculation is not the same as the original paper, and a similar measure is used as a proxy. The two cases in which this occurs are the following: Firstly, data could not be collected on the level of investment in order to calculate the CMA factor. As a proxy, return on capital employed is used, because companies that invest less will be earning a higher return on their investments, assuming that the available projects suffer from diminishing marginal returns. Therefore, the CMA factor instead compares companies with high ROIC to low. Secondly, the volatility factor is calculated in a similar fashion to other factors, this is to ensure that all predictors have the same units. It is important that all factor variables are measured in percentage return because the shrinkage methods implemented in section 4 shrink based on absolute or squared coefficient values. Therefore, if each variable was measured in different units then these methods would be biased toward certain predictors.

Now that factor returns have been constructed, it may provide useful context to briefly investigate these variables. Figure 2 shows the returns of each factor relative to the market since 2010, while table 2 specifies the risk adjusted performance of each factor. Over the 1995-2018 period that the data spans, the market had a Sharpe ratio of just over 0.2. Over the same period, only the size and low volatility factor did not outperform the market, meanwhile the best performing factors on a risk adjusted basis over the period were those measuring quality, namely low investment and profitability. Note also the strong performance of the low investment factor, this is most likely due to the positive relationship between ROIC and equity returns leading to a circular relationship between the two. This issue is acknowledged but the low investment factor remains calculated in this manner, as later regressions will show no distorting improvement in model fit due to this. Researchers looking to build upon results from this thesis are encouraged to consider using asset growth as a measure of low investment, conditional on data availability.

Figure 2: Factor Performance Relative to Market

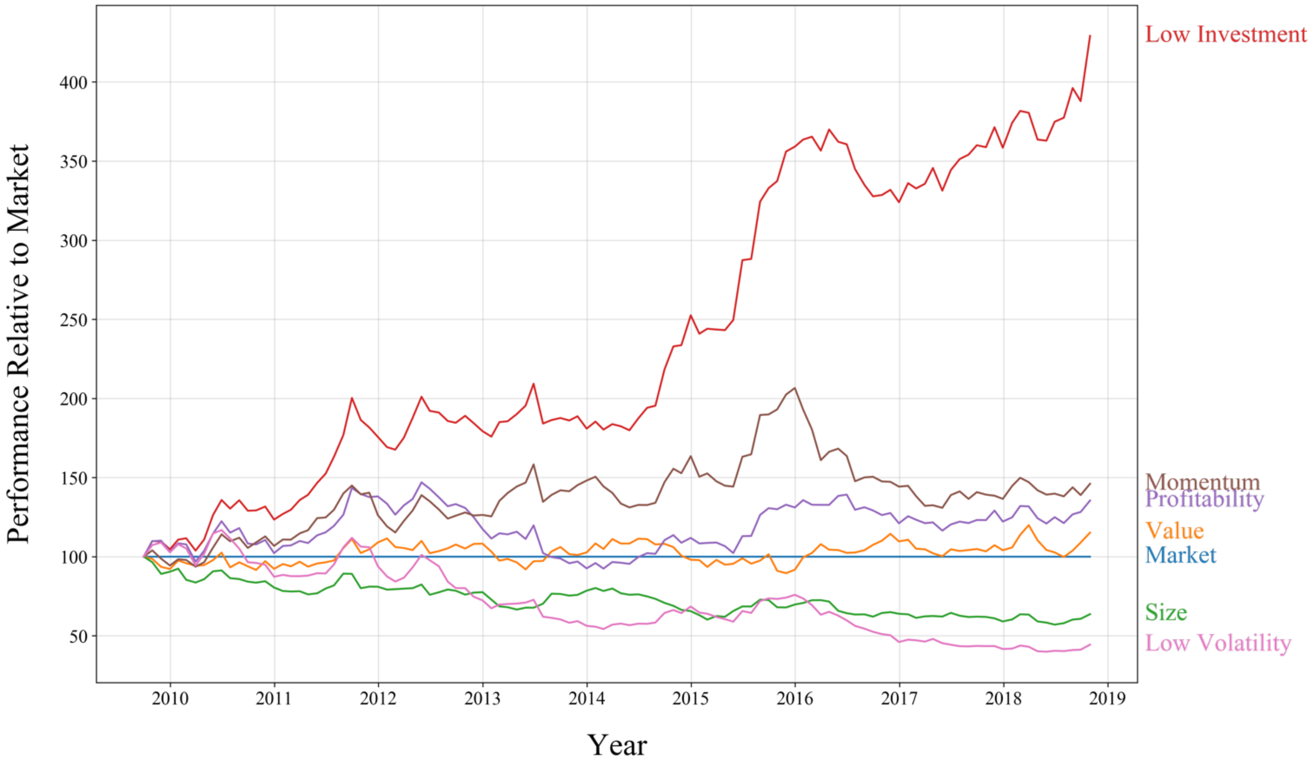


Table 2: Sharpe Ratios of UK Equity Factors

	Sharpe Ratio
Market	0.207
Value	0.345
Size	0.108
Low Investment	0.757
Profitability	0.409
Momentum	0.312
Low Volatility	0.017

Finally, each higher order polynomial is calculated by raising the gross factor return to the respective power, converting to net return after the calculation is done. For example, if SMB return is 2%.

$$SMB = 0.02 \tag{13}$$

Then the higher order polynomials as in equation 14.

$$SMB^2 = (1 + 0.02)^2 - 1 \tag{14}$$

$$SMB^2 = 0.0404 \tag{15}$$

3.3. Defining Predictor Sets

As mentioned, there are two main avenues through which nonlinearity can be included in a factor model, the first is by creating a set of nonlinear independent variables by including higher order polynomials and interaction terms. These nonlinear predictor sets are generated using as a base the models discussed in section 2.2, by including each variable up to the cubic polynomial as well as including interaction terms between these variables. The linear forms of each model are specified in equations 16 to 20 respectively, meanwhile a full list of the nonlinear predictor sets can be found in the appendix. Finally, in addition to the common factor models, a combined model is constructed consisting of the union of all the above models in order to, among other reasons, investigate the extent to which extensions of the more basic models are capturing the same return variation.

FF (1993) Three-Factor Model

$$r_{i,t} - r_{f,t} = \alpha + \beta_1 MKT_t + \beta_2 HML_t + \beta_3 SMB_t + u_{i,t} \tag{16}$$

Carhart (1997) Four-Factor Model

$$r_{i,t} - r_{f,t} = \alpha + \beta_1 MKT_t + \beta_2 HML_t + \beta_3 SMB_t + \beta_4 UMD_t + u_{i,t} \tag{17}$$

FF (2015) Five-Factor Model

$$r_{i,t} - r_{f,t} = \alpha + \beta_1 MKT_t + \beta_2 HML_t + \beta_3 SMB_t + \beta_5 CMA_t + \beta_6 RMW_t + u_{i,t} \tag{18}$$

Ang (2006) FF Three-Factor Model with Volatility

$$r_{i,t} - r_{f,t} = \alpha + \beta_1 MKT_t + \beta_2 HML_t + \beta_3 SMB_t + \beta_4 VOL_t + u_{i,t} \tag{19}$$

Combined Model

$$r_{i,t} - r_{f,t} = \begin{array}{c} \alpha + \beta_1 MKT_t + \beta_2 HML_t + \beta_3 SMB_t + \\ \beta_5 CMA_t + \beta_6 RMW_t + \beta_7 UMD_t + \beta_8 VOL_t + u_{i,t} \end{array} \tag{20}$$

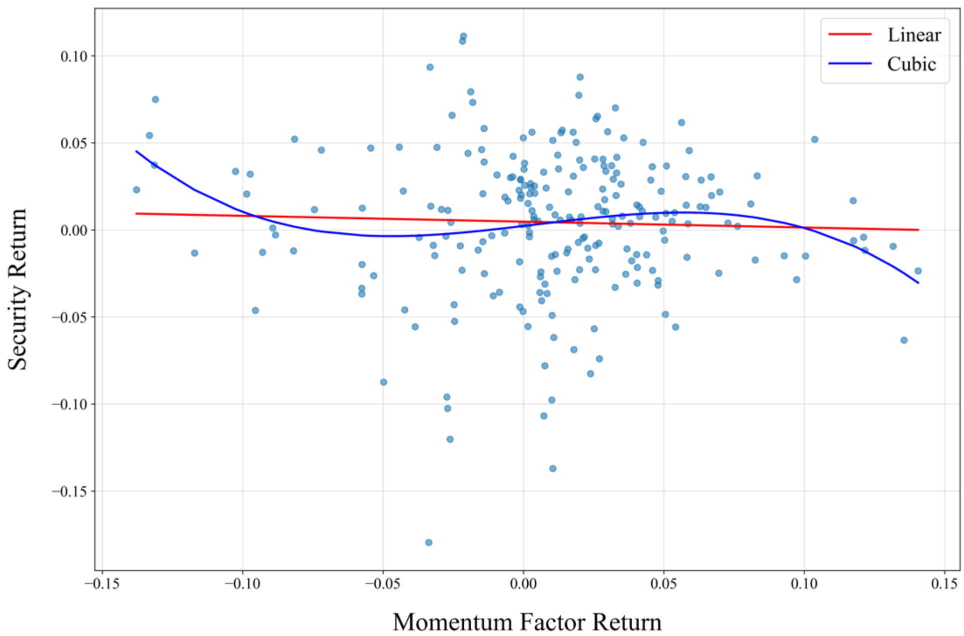
3.4. Evidence of Nonlinearity

The final task in this section is to provide a case for the potential nonlinearity in UK equity risk premia. A brief investigation into the relationship between various independent variables and security returns indicates that there is a case for the nonlinear modelling of equity factors. This section outlines this rationale via three avenues; looking at these relationships graphically, statistically testing the significance of nonlinear predictors, and using a correlation matrix of various predictors.

Firstly, the graphical approach. Figures 3 and 4 show the relationship between security return and the size and momentum factors respectively as examples of the nonlinearity that is present. In both cases the argument can be made that a linear model would fail to effectively fit the respective relationship, or at least that a nonlinear model would be expected to perform better.

Figure 3: Size (SMB) Factor vs Security Return

Figure 4: Momentum (UMD) Factor vs Security Return



The shape of figures 3 and 4 suggest that the relationship between factor returns and security returns is conditional upon the factor return itself, meaning that factor returns at the tails of the distribution have a different relationship with factors than those toward the centre. This highlights that the security return – factor return relationship does not follow the traditional pattern during unprecedented times and anomalous market situations.

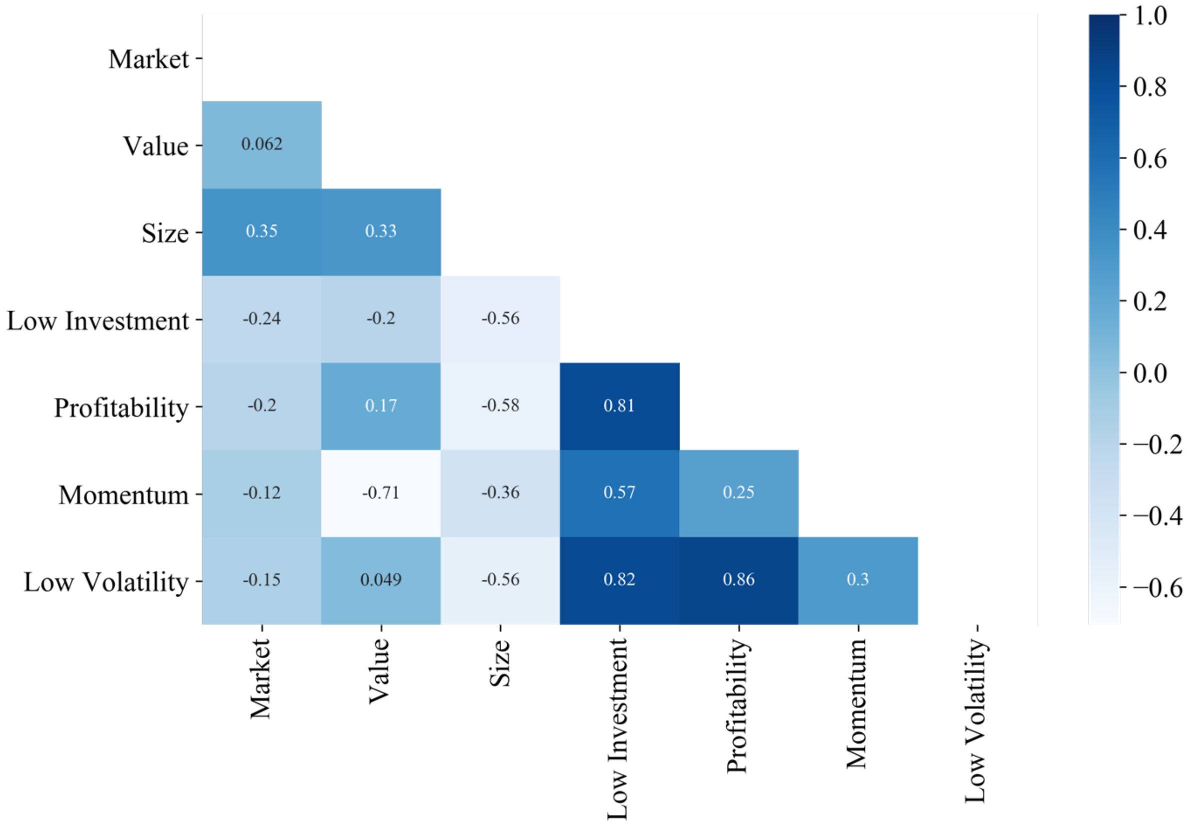
Secondly, quadratic and cubic polynomials of each factor are statistically significant when included in a polynomial regression model to predict excess returns. In fact, in the case of the SMB factor these higher order polynomials are more significant than the linear form of the variable. Table 3 shows a polynomial regression of the FF five factor model on returns over the first 80% of the dataset.

Table 3: Polynomial Regression Coefficients of Combined Model

	Coefficient	*P-Value*
Constant	0.022	0.000
Market	87.858	0.000
$Market^2$	-86.102	0.000
$Market^3$	28.132	0.000
Size	10.757	0.058
$Size^2$	-13.925	0.011
$Size^3$	4.446	0.012
Value	27.796	0.000
$Value^2$	-27.796	0.000
$Value^3$	9.311	0.000
Profitability	-105.435	0.000
$Profitability^2$	101.897	0.000
$Profitability^3$	-34.454	0.000
Low Investment	17.587	0.007
$Low\ Investment^2$	-23.285	0.000
$Low\ Investment^3$	7.276	0.000
$Value \times Size$	0.161	0.563
$Low\ Investment \times Profitability$	4.134	0.000
$Low\ Investment \times Size$	3.014	0.000

Finally, a correlation matrix can be formed to picture the strength of the relationship between different factor returns. Figure 5 suggests that there may be interaction effects between them, supporting some of the relationships that are known to exist between factors. For example, small companies that have a higher exposure to the SMB factor also tend to be better value for money, as defined by a higher exposure to the HML factor.

Figure 5: Factor Returns Correlation Matrix



4. Methodology

This section discusses the models that are used, as well as how each is fitted and validated. Section 4.1 describes the different models that are to be compared, while sections 4.2 and 4.3 outline the validation procedure used to fit the models and optimise their tuning parameters.

4.1. Models Estimated

The goal of this thesis is to evaluate if allowing for nonlinearities in the relationship between common equity factor returns and security returns improves upon the standard linear regression. Further to the creation of polynomial predictors, various different linear and nonlinear models can be applied to these predictors. This section summarises the specifications of each model applied in this thesis.

4.1.1. Ordinary Least Squares

Before investigating the performance of non-linear models, it is necessary to first estimate the linear model in order to provide a benchmark for their performance. The Ordinary Least Squares (OLS) model aims to minimise the sum of squared residuals (SSR) as defined in equation 21.

$$SSR = \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{k} x_{j,i}\beta_j \right)^2 \tag{21}$$

Given that all residuals are weighted the same, large outliers can have a disproportionate effect on the performance of the model and its predictions, as outlined in section 3.4.

The OLS model is easily extended to allow for nonlinearity by including higher power independent variables and interaction terms in the model. The polynomial regression applies OLS to the nonlinear predictor set defined in section 3.3 using the same objective function. The rationale for this model is that by permitting a nonlinear relationship between predictors and returns, the model will be a better fit. Lastly, in terms of the order of the polynomial, no strategic process was used to choose a cubic power, but is chosen largely due to statistical conventions used in running polynomial regressions. However, the argument could be made that a cubic polynomial regression be used instead of a simpler quadratic model in order to allow for additional nonlinearity, meanwhile a fourth power predictor set runs the risk of overfitting.

4.1.2. LASSO Regression

LASSO stands for "Least Absolute Shrinkage and Selection Operator". It is a form of penalised regression in which a shrinkage factor attempts to limit the number of predictors incorporated into the model. The same predictor set is used as the polynomial regression, ensuring that this model can incorporate nonlinearities as well. The LASSO regression implements L1 regularisation, which means that it includes a penalisation term which applies to the sum of the absolute values of predictor coefficients. Mathematically, instead of minimising the SSR, the optimisation aims to minimise equation 22.

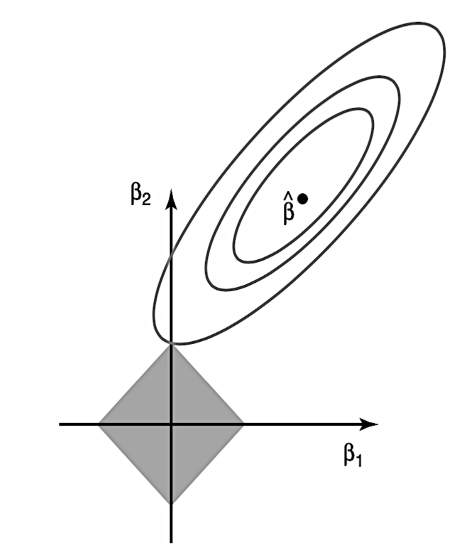
$$\sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{k} x_{j,i}\beta_j \right)^2 + \lambda \sum_{j=1}^{k} |\beta_j| \tag{22}$$

The shrinkage term, λ, determines the severity with which additional and larger predictors are penalised. A higher value of λ will limit the size of the predictor set more strongly than lower values, and so the choice of λ is important, as explored in section 4.3.

Shrinkage methods fit a model whose coefficients lie at the tangency of the contours of the error function and the constraint region. In the case of the LASSO model the constraint region is square shaped with its centre at the origin as shown in figure 6.

Figure 6: LASSO Constrained Optimisation Problem

(Source: James et al., 2013)



As a consequence of this, the LASSO model has a tendency to reduce some coefficients to zero, while not altering others.

4.1.3. Ridge Regression

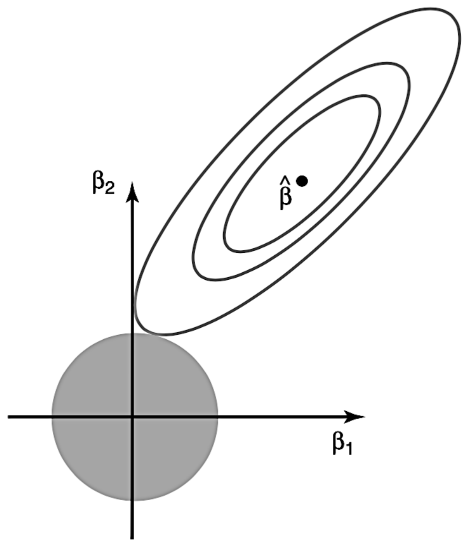
The ridge regression also falls into the category of penalised regressions. It is similar to the LASSO, but instead of using L1 regularisation to constrain the number of predictors it implements L2 regularisation. L2 regularisation means that instead of considering the sum of the absolute values of coefficients, the Ridge regression penalises the sum of the squared coefficients. This results in an optimisation problem which aims to minimise equation 23.

$$\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{k} x_{j,i}\beta_j\right)^2 + \lambda\sum_{j=1}^{k}\left(\beta_j\right)^2 \tag{23}$$

In the case of the ridge regression, the constraint region is circular due to the fact that coefficients are squared in the optimisation problem in equation 23. As a result, in contrast to the LASSO model, the ridge model will reduce all coefficients by a small amount as opposed to some coefficients to zero, depicted graphically in figure 7 (James et al., 2013).

Figure 7: Ridge Constrained Optimisation Problem

(Source: James et al., 2013)

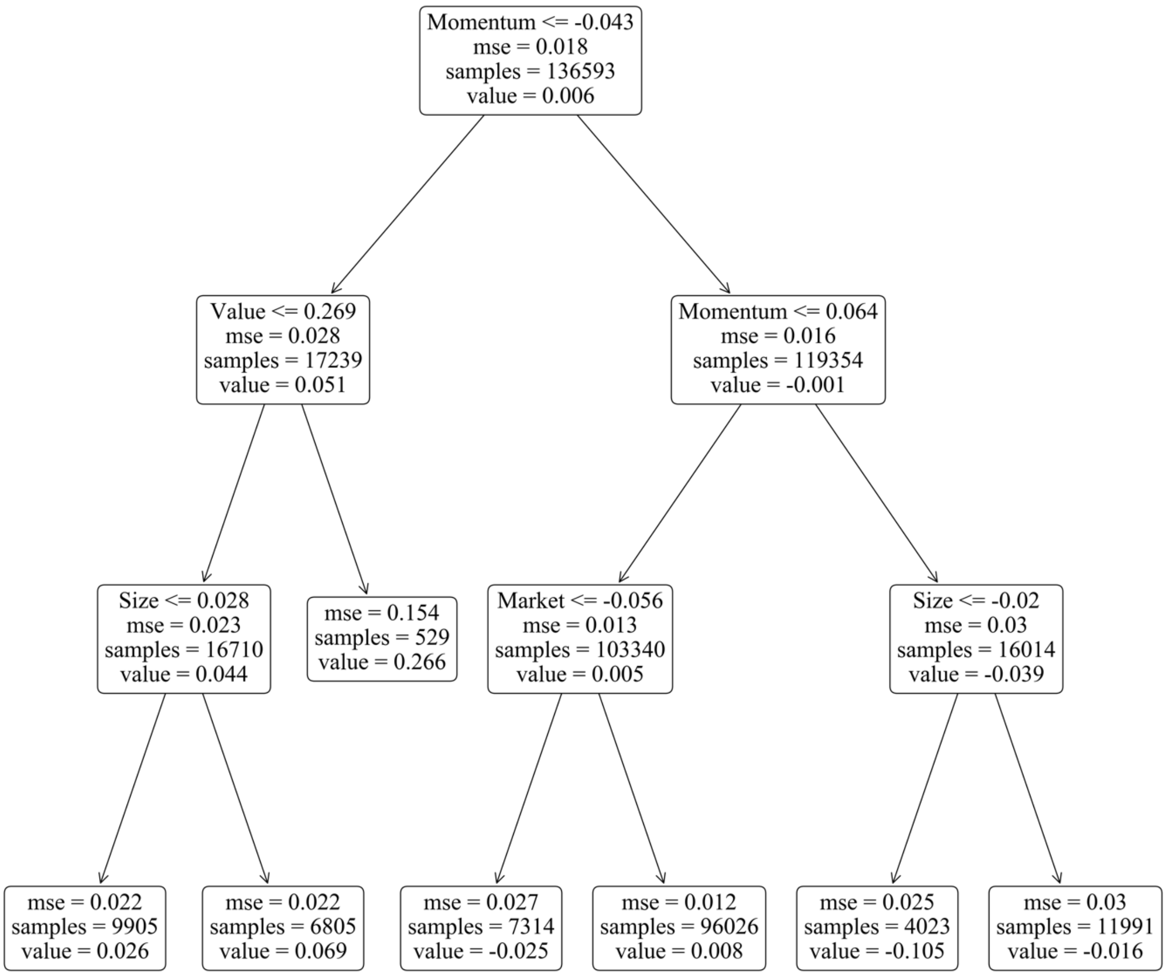


4.1.4. Regression Tree

The final model used is a regression tree, which separates the data into a number of different, non-overlapping regions according to its characteristics. For each group it then makes a prediction as its mean value. The model is trained via a process of recursive binary splitting, in which the algorithm aims to split the data into two groups in such a way as to minimise the SSR as defined in equation 21. The same splitting approach is then applied to each of the subgroups into which the data has been split. This process is then repeated until no more groups are required or, more commonly, when a predefined limit to the size of the tree is reached. The process is said to be "greedy", because it only optimises with respect to the current node and does not consider future groupings, mainly because doing so is too computationally intensive.

An example of a regression tree can be found in Figure 8, which shows the fitted tree for the Carhart four factor model applied to the UK dataset used in this thesis. Note that predictors can be repeated within a tree, for example the SMB factor is repeated on the internal nodes of the left-hand side of Figure 8.

Figure 8: Regression Tree Output of Carhart Four Factor Model

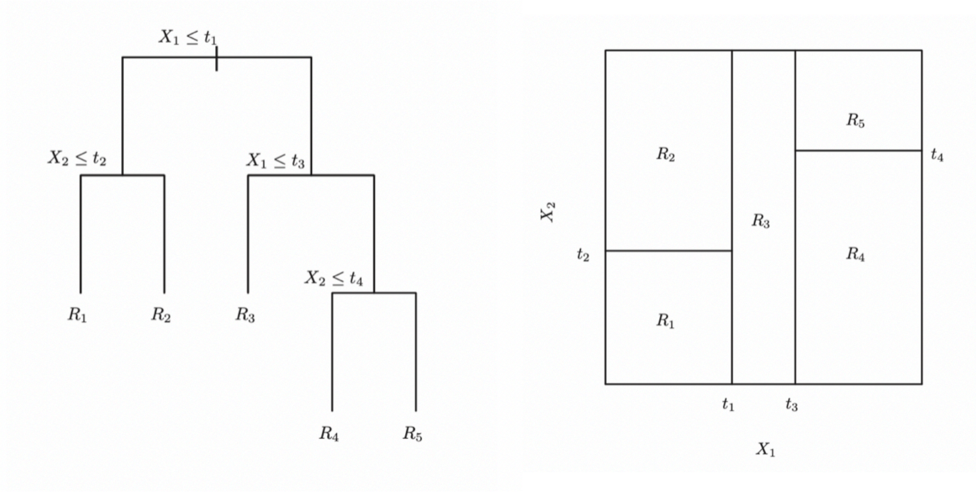


As mentioned, when fitting a regression tree model, it is required to choose a maximum tree depth. This essentially asks how many different groups the data are to be split into. The value of this parameter is chosen via a validation process outlined in section 4.3.

Regression trees are different to the rest of the models in that they are nonparametric, meaning that no functional form is implied in the relationship between predictors and outcomes when training the model. Consequently, trees can be rather intuitive as the groupings are easily explained by the decision nodes and predictor importance coefficients. Figure 9 shows a simple regression tree and the corresponding regions of the two-dimensional feature space.

Figure 9: Regions of a Regression Tree

(Source: James et al., 2013)



4.2. Model Training

Each of the models outlined in the previous section is fitted on both the linear and nonlinear predictor sets. Traditionally, a process of k-fold cross validation is used to train models and evaluate their true performance. K-fold cross validation involves dividing the dataset into k segments, or "folds". One by one each fold is temporarily removed from the dataset. The model is then trained on the other $k - 1$ folds and tested on the fold that was left out. Once each fold has been tested, an average performance can be taken over all folds and a fair evaluation of the model's performance made.

The out-of-sample performance of each model when trained via a validation process can be gauged by the Mean Squared Error (MSE) of the model's predictions. MSE is defined in equation 24 and is a common error measure in both statistics and machine learning.

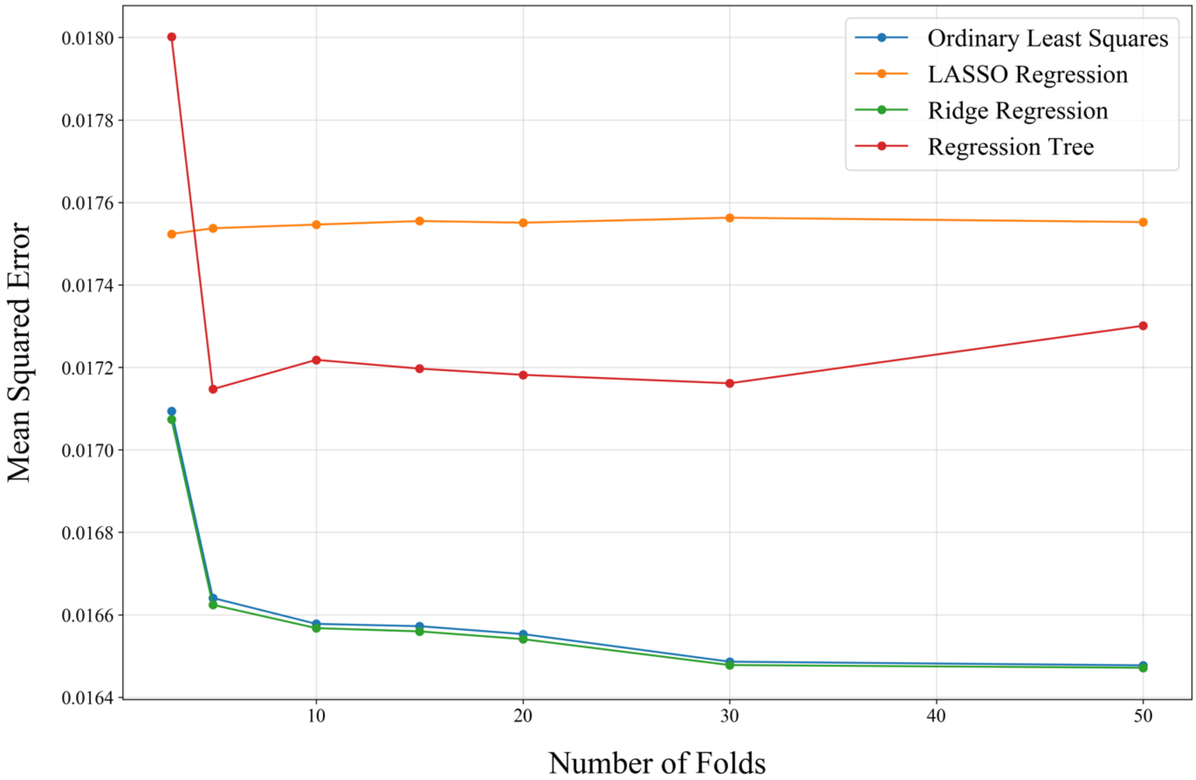
$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{24}$$

Consequently, the performance as measured by k-fold cross validation is defined in equation 25.

$$CV_{(k)} = \frac{1}{K}\sum_{k=1}^{K} MSE_k \tag{25}$$

Figure 10 shows the performance as measured by MSE in modelling the FF five factor relationship over different numbers of folds.

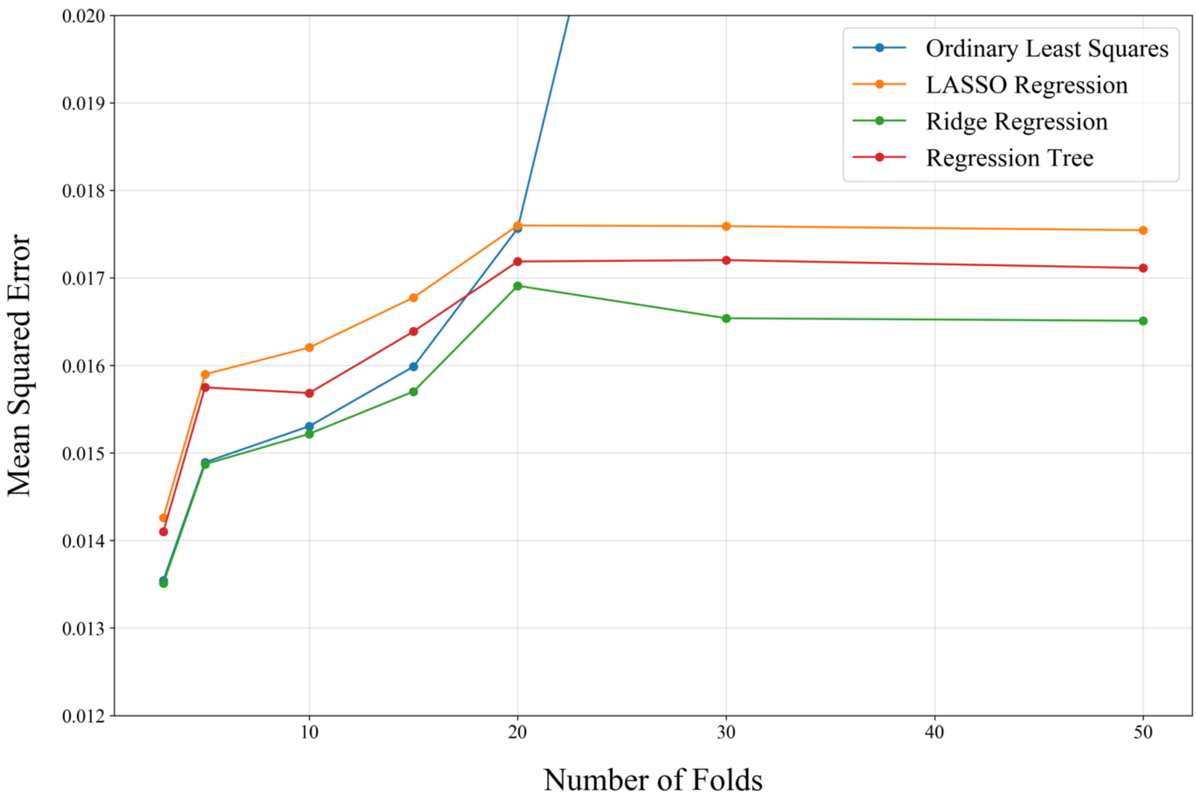
Figure 10: Plot of average MSE vs number of folds used in k-fold CV



K-fold cross validation is commonly used in situations where the ordering of data does not matter. However, it can cause issues when applied to time series data given that it does not account for the intertemporal relation between different folds. When using this method, one is essentially trying to predict the past given what happened in the future, which not only lacks intuition but may also generate distorted estimates of out-of-sample performance.

Consequently, an extension of the k-fold method is used instead. Rolling origin (or "walk forward") evaluation (Tashman, 2000) uses the same approach as above, however, instead of randomly removing a fold, the folds are removed chronologically such that each model is trained on historical data and tested on future data. This process is then repeated sequentially so as to maintain the temporal ordering of the data. Figure 11 shows the performance of each model as a function of the number of folds.

Figure 11: Plot of average MSE vs number of folds in rolling origin CV



The average results from rolling origin evaluation are fairly intuitive. As the number of folds increases, the time horizon for out-of-sample testing becomes shorter and shorter, causing randomness to become a larger and larger factor in determining a security's return. Given models of equity factors are inherently statistical models that depend on laws such as that of large numbers and expectation, the models will fit much less well across shorter time horizons. This is what leads to the increase in MSE in Figure 11 as the number of folds increases, before levelling out beyond 20. On a separate note, the poor performance of the OLS model using cubic predictors occurs due to the lack of penalisation on higher order coefficients, leading to incorrect predictions being made based on erroneous historical relationships. This is not included in the figure to allow for a more specific comparison of the models that perform well.

In order to strike a balance of performance between k-fold cross validation and rolling origin evaluation, five folds are used in constructing models. This is done by taking an average of the five models' coefficients to create a validated model whose out-of-sample performance is expected to be in line with that of figure 11.

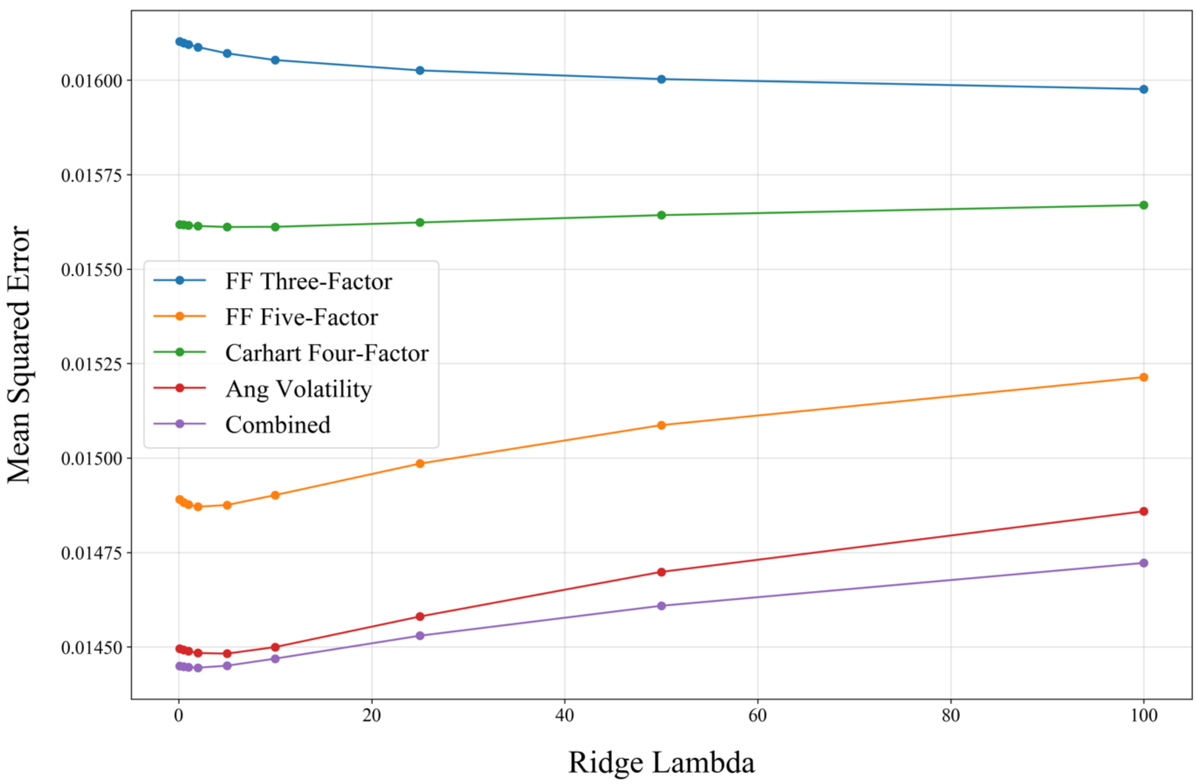
4.3. Tuning Parameter Optimisation

There are three tuning parameters that enable the respective models to be optimised. The first two are the λ values associated with the ridge and LASSO regressions, and the third is the maximum tree depth of the regression tree model. When optimising each parameter, the goal is to minimise the out-of-sample mean squared error (MSE) while avoiding overfitting, known as the "bias-variance tradeoff".

Firstly, the λ parameter associated with the LASSO regression does not impact the coefficients of the model. For any $\lambda > 0$ the fitted LASSO regression predicts a constant return only, reducing all other coefficients to 0. This leaves the choice of λ arbitrary, and so for simplicity the λ value of the ridge regression is chosen.

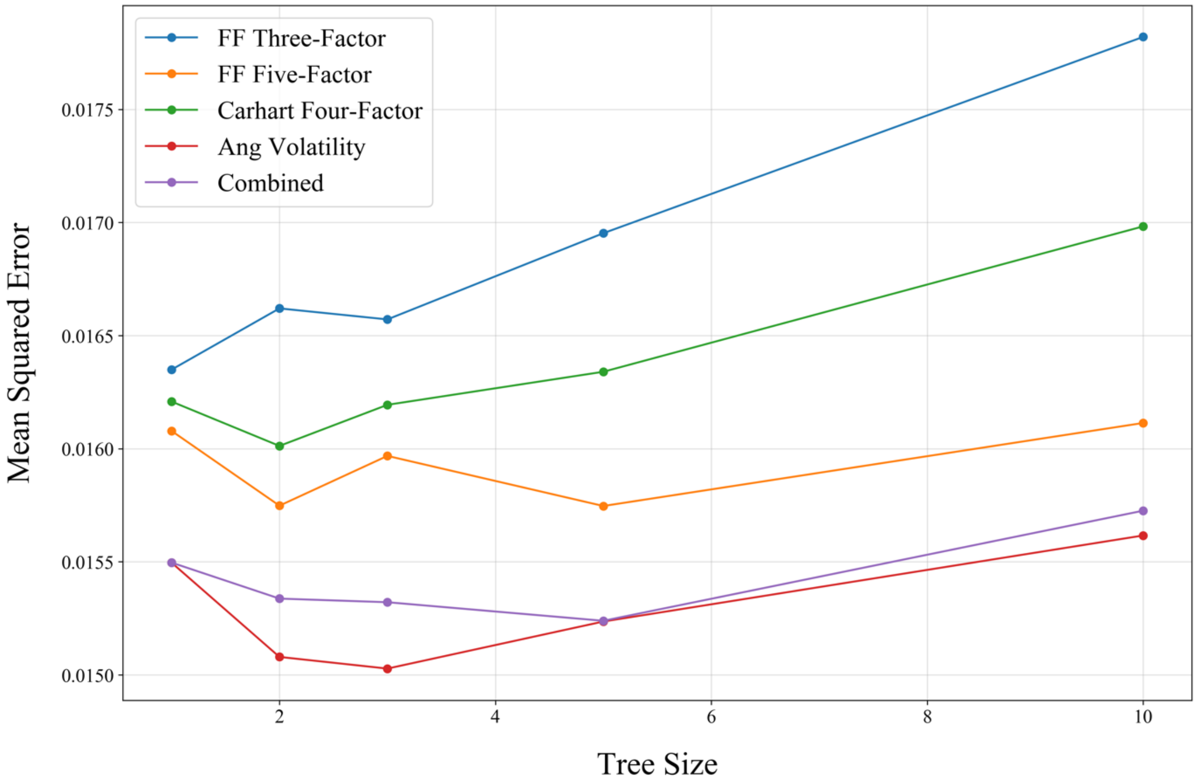
Secondly, the optimisation process for the λ parameter of the ridge regression suggests that only a small penalty on coefficient size is optimal. Figure 12 shows the MSE of each model at different values of λ. While in the FF three-factor and Carhart four-factor models there is not much change in the MSE, the other three models appear to perform worse as λ increases. As a result of this the low but not insignificant value of two is used for the λ parameter.

Figure 12: MSE vs Ridge Regression λ



Finally, the maximum tree depth associated with the regression tree model refers to largest hierarchical length that is permitted. A higher value will make the model more specific but at the same time more likely to be overfitted to the training data. Figure 13 shows how each model performs over different tree depths, indicating that small tree size is likely to be optimal, with a maximum depth of three being chosen for the fitted models.

Figure 13: MSE vs Maximum Tree Depth



5. Results

This section summarises the results and discoveries of the thesis. Results on model performance are broken down by the predictor set that is used, hence performance tables will be presented in pairs, one each for the linear and polynomial predictor sets. Model performance is shown in terms of average out-of-sample R^2, which as expected are low due to the high degree of noise present in financial market data.

Firstly, a note on out-of-sample R^2 values. When testing model performance on unseen data, R^2 values can be negative because the models have the potential to be fitted on data that has a significantly different relationship. This means that the trained model could perform worse than a random or constant prediction if the training and testing data are too dissimilar. By contrast, in-sample R^2 values can only range between 0 and 1 because when optimising training performance (as all models do) there is always the backup option of a constant prediction. Essentially, a model that works well on training data may be worse than a random guess when applied to unseen data, especially in in the case of security returns where the degree of noise is so high. With this in mind, positive R^2 values are highlighted in **bold** for clarity.

5.1. Model Performance

When comparing the performance of the linear and cubic predictor sets against each other, the addition of nonlinearities in the independent variables does not appear to lead to any improvement in goodness of fit. Tables 4 and 5 show the performance of each model according to rolling origin evaluation (Tashman, 1997). In almost every case, the performance of a model is better when using the linear predictor set than its cubic counterpart. This may suggest that the inclusion of additional variables does more harm to the models in terms of complexity than it adds in goodness of fit. The linear models, being simpler, appear to be more versatile to unseen data and less susceptible to overfitting. As a result of this, the analysis that follows in this section is based on the linear specifications of each model unless otherwise stated.

Table 4: Average R^2 of Models using Linear Predictors

	OLS	*LASSO*	*Ridge*	*Tree*
FF Three-Factor	-0.014	-0.001	-0.013	-0.068
FF Five-Factor	**0.033**	-0.001	**0.035**	-0.012
Carhart Four-Factor	**0.005**	-0.001	**0.006**	-0.015
Ang Volatility	**0.056**	-0.001	**0.057**	**0.027**
Combined	**0.06**	-0.001	**0.061**	**0.008**

Table 5: Average R^2 of Models using Polynomial Predictors

	OLS	*LASSO*	*Ridge*	*Tree*
FF Three-Factor	-0.042	-0.001	-0.026	-0.026
FF Five-Factor	-0.658	-0.001	**0.024**	**0.004**
Carhart Four-Factor	-0.598	-0.001	**0.009**	**0.006**
Ang Volatility	-0.333	-0.001	**0.057**	**0.017**
Combined	-0.426	-0.001	**0.057**	**0.008**

Although the linear predictor set appears more performant, a comparison of each model in table 4 suggests that despite the inclusion of nonlinear predictor sets not improving goodness of fit, the application of different models to the linear predictor set can improve fit. Results from the ridge regression in table 4 consistently outperform that of the linear OLS. A potential reason for this could be the susceptibility of OLS to outliers which increase the estimated sensitivity of security returns to each factor, a characteristic that is counteracted by the coefficient penalisation term of the ridge regression.

This supports the argument made in section 3.4 that those security returns that lie on the tails of the distribution do not have the same relationship to factor returns as those in the centre.

In addition to measuring model performance using R^2, the extent to which a factor model captures all return variation can be evaluated via an analysis of the regression intercept. FF (1993) showed that the estimated intercept in their three-factor model was statistically insignificant from 0, leading to the conclusion that their model of security returns was a complete one given that there was no return unaccounted for. This thesis applies the same analysis to evaluate the completeness of each factor model by testing the intercept of the OLS model. Table 6 shows the p-value of the OLS intercept coefficient using both linear and cubic predictor sets for each model. Of all the models tested with a linear specification, the FF three-factor model, the Ang volatility model and the combined factor model are the only two which have an insignificant intercept at the 10% level, and hence are potentially complete factor models.

Table 6: P-Values of OLS Intercept

	Linear	*Polynomial*
FF Three-Factor	0.122	0.000
FF Five-Factor	0.000	0.000
Carhart Four-Factor	0.000	0.000
Ang Volatility	0.092	0.000
Combined	0.108	0.142

5.2. Model Coefficients

In addition to looking at the intercepts, an investigation is made into the variable coefficients of each model. The model results supplied in this section relate to the linear form of each model. This is both because model performance when using linear specification was better, and because linear coefficients allow for much easier interpretation. Results using the cubic predictor sets can be found in appendix C.

In addition to agreeing with some of the common research from the US, the plain vanilla linear model suggests some relationships that are in contrast to the standard literature. Table 7 shows the regression coefficients of the combined linear factor model using OLS. The insignificance of the profitability and low investment factors variables suggests that another factor is capturing the same variation. Upon further investigation this appears to be the volatility factor, as shown by regressions in appendix C. Furthermore, the market and value factors are as we would expect them to be, while the remaining factors appear in contrast to standard theory. There are three potential reasons for this. Firstly, most equity factor models are fitted on US data and hence will find a different relationship to that of this thesis. Secondly, the timeframe of the data in this thesis is much more modern than those of traditional models, with this thesis using data from 1995 whereas most traditional models use data that ends around this period. Finally, the calculation of the volatility factor is not the same as in FF's original research, in order to standardise the units in regressions.

Table 7: Linear OLS Coefficients

	Coefficient	*P-Value*
Intercept	0.003	0.108
Market	0.262	0.000
Size	-0.453	0.000
Value	0.336	0.000
Profitability	-0.240	0.110
Low Investment	0.175	0.176
Momentum	-0.109	0.000
Low Volatility	-0.595	0.000

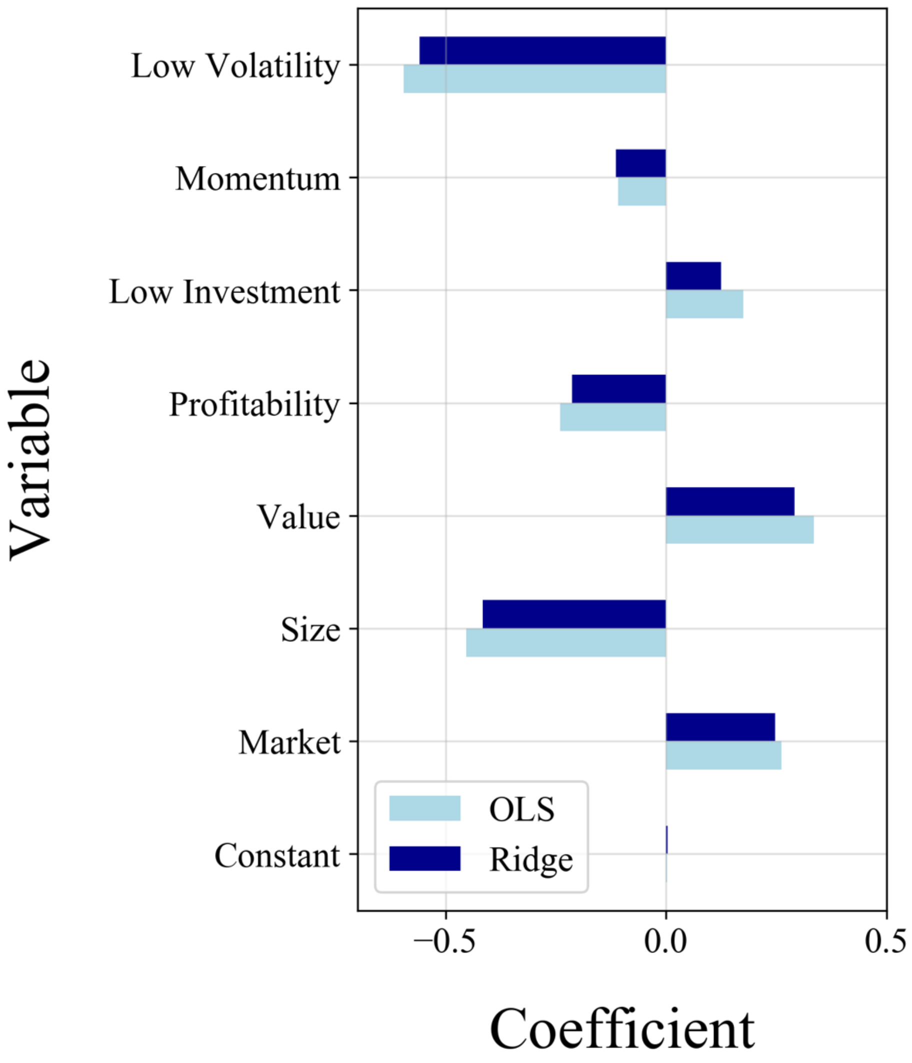
The two shrinkage methods employ coefficient dampening very differently. The coefficients for both models are shown in table 8. The coefficients of the ridge model broadly coincide with that of OLS, whereas the LASSO model predicts a constant return in each period, reducing all coefficients to zero. This prediction of 0.7% monthly return amounts to an expected return of 8.7% per annum. Meanwhile, the ridge regression coefficients are shown in the right-hand column of table 8, comparing these to the standard polynomial regression with no shrinkage we can see which variables are of most or least importance.

Table 8: LASSO and Ridge Regression Coefficients

	LASSO	*Ridge*
Intercept	0.007	0.004
Market	0	0.248
Size	0	-0.416
Value	0	0.291
Profitability	0	-0.213
Low Investment	0	0.125
Momentum	0	-0.114
Low Volatility	0	-0.559

A more analytical comparison of the coefficients of the OLS and ridge models is conducted in order to identify those coefficients that are more or less important, judging by the amount the coefficient is reduced. Figure 14 shows how the regression coefficients of OLS and ridge compare. As expected, most of the coefficients of the ridge model are lower in absolute terms than those of the OLS model, with the exception of the UMD factor which is slightly higher.

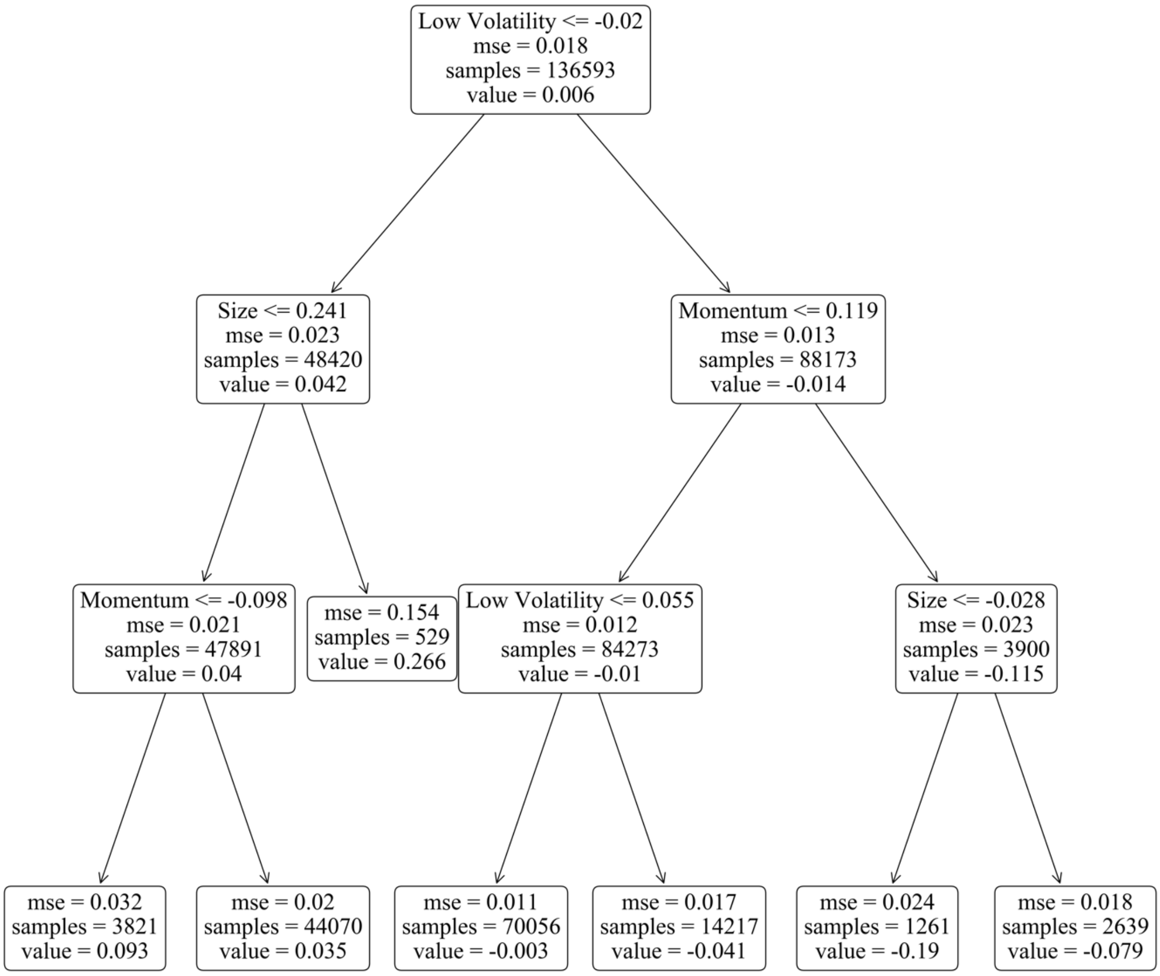
Figure 14: Coefficient Differences Between Ridge and OLS



One key piece of information must be kept in mind when using shrinkage models. We can use these models for factor regressions because all predictors are in the same units, so there is no bias toward shrinking any predictor. However, when used for prediction these models lose a significant amount of value if variables are not standardized, because the independent variables whose coefficients they are shrinking are not all of the same units. So, those predictors whose units are relatively small compared to the units of the dependent variable (percentage return) will be penalized more harshly.

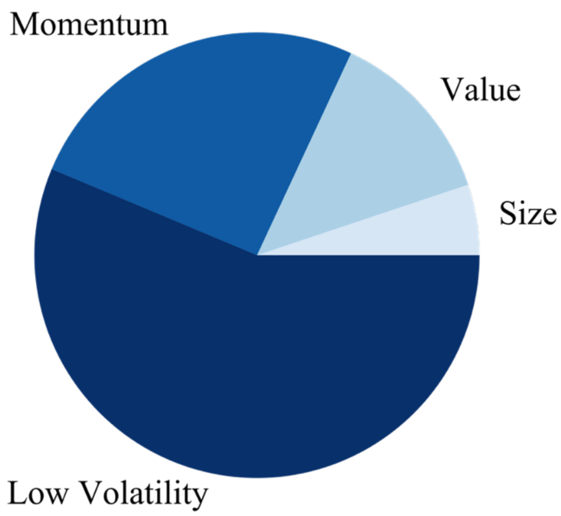
Finally, an analysis of the fitted regression tree is conducted. Figure 15 gives us a valuable insight into how the fitted regression tree is structured, as well as the coefficients that are more useful in splitting data.

Figure 15: Regression Tree of Linear Combined



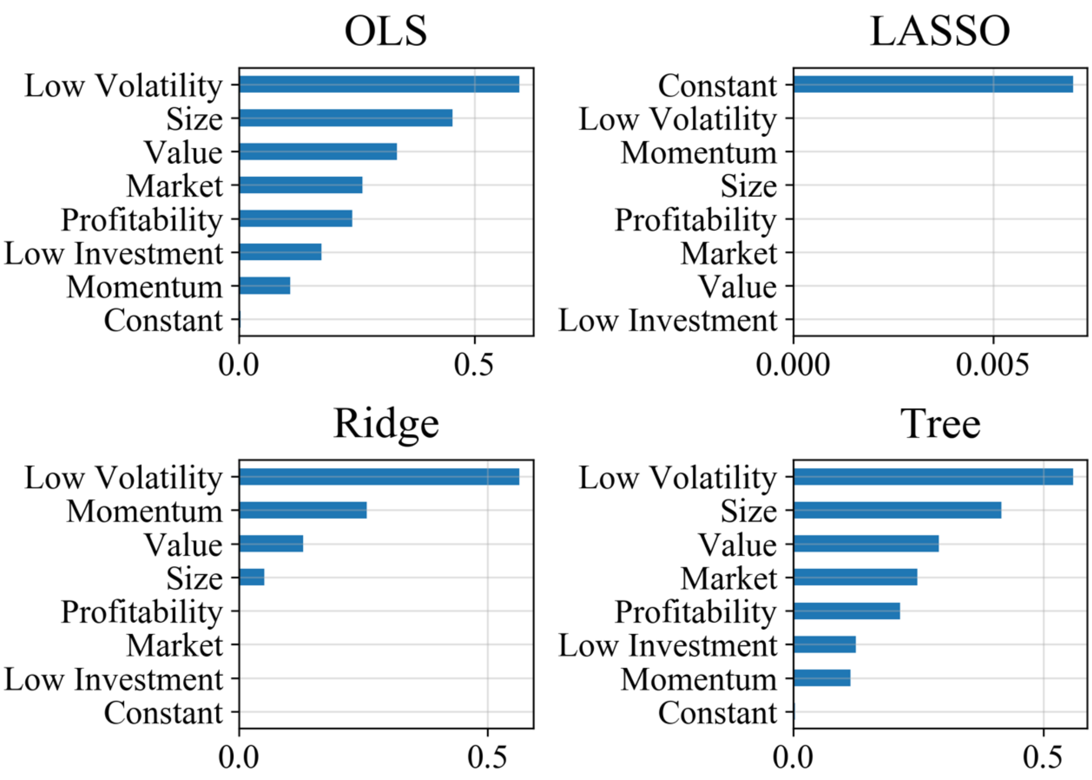
In addition to viewing the tree structure the individual importance of each predictor can be calculated. Put simply this importance value is defined as the amount by which the feature reduces the MSE of the model, normalised such that all values sum to 1. Figure 16 shows the relative importance of each variable in the tree model. As has been the case in the other regression models, volatility plays a large part in fitting the models, as do size and value. This supports the argument made earlier that the FF three-factor and Ang volatility models may be complete models of factor returns.

Figure 16: Feature Importances in Linear Combined Model



From a more holistic viewpoint, the models show a fairly similar value ordering toward the variables in terms of the importance placed on each predictor. In a similar fashion to Gu, Kelly and Xiu (2018), for each statistical model the importance of each variable is calculated, and the results show in figure 17. Firstly, the data show the significant role played by the volatility factor when fitting the models, it ranks highest among all but the LASSO model. In addition to this, the size and value factors from the FF three-factor model also have unanimous significance across almost all models.

Figure 17: Variable Importance by Model



In summary, the results shown in this section suggest that not only can different statistical models improve on the linear OLS specification, but that when the linear model is applied to a more recent dataset, and to the UK instead of the US, the resultant factor return relationships may be different. Of course, further analysis would need to be conducted, but this could have significant importance for local investors and traders who may be using literature specific to the United States to make investment decisions.

5.3. Limitations

As a corollary to the aforementioned results, it is recognized that in any research the results come with practical limitations, this thesis being no different. The following is a brief discussion of the limitations faced in this thesis and the potential consequences on the aforementioned results.

Firstly, data covering the UK market is not as rich or lengthy as that of the US, which has been the basis for most research on risk premia. Despite this issue, a motivation for this thesis is to investigate what information can be extracted from the data that is available, while acknowledging its limitations.

Secondly, the results presented may not be of the highest economic significance given that the models fitted do not account for practical issues such as that of trading costs. Hence it is admitted that it may not be practically possible to act on the positive performance of the fitted models, and in fact this could be a reason for a positive R^2 in the first place.

Thirdly, the extent to which underlying variables indicate differences in securities could be limited by practical issues. For example, the profitability factor relies on information about the operating profit margin of a company. However, this measure is largely down to accounting standards and decisions around things like depreciation and amortisation. Consequently instead of profitability, it could be better to use a measure of cashflow instead.

Finally, it is true that machine learning methods and other nonlinear statistical models are easier to overfit. The validation processes implemented aim at solving this, but one must still be aware of the bias-variance trade-off, specifically when using nonlinear models.

6. Making Predictions

So far, this thesis has looked at how a variety of nonlinear models compare to the linear regression in fitting a relationship between security returns and common factor models. However, nonlinear statistics and machine learning techniques are also designed to make predictions. Therefore, this section will briefly investigate the comparative efficacy of these models in predicting future security returns instead of modelling them contemporaneously. Understandably the performance here is weaker, as any significant results are likely to be arbitraged away by the millions of market participants looking for trading opportunities each day. When used in prediction, contemporaneous factor returns cannot be used as predictors because only the factor returns for the previous period are available which are unlikely to have much predictive power. Therefore, the underlying variables used to construct factor returns are regressed on instead. For example, in place of the HML factor, the underlying book-to-market value is regressed instead.

Of the models tested, regression trees appear to be particularly effective in return prediction as opposed to return decomposition. Tables 9 and 10 show the regression tree to be the only model with any predictive power, posting a 0.4% R^2 when using both the linear and nonlinear specifications. This result is supported by Verbiest (2011) who finds strong out-of-sample predictive power in regression tree models when a seven-factor setup similar to that of the combined model that this thesis uses. Meanwhile, when fitting each model on the nonlinear set of variables, the FF (2014) five-factor model leads the regression models to perform extremely poorly. Table 10 shows extremely negative R^2 values for the OLS, ridge and LASSO models. The most likely reason for this is that the profit margin data has some anomalous entries which lead to a poorly fitted model that fails to predict returns accurately. This may be supported by the relatively strong performance of the regression tree model which does not include the profitability metric at any decision node, as shown in appendix D.

Table 9: Prediction Performance of Linear Specifications

	OLS	*LASSO*	*Ridge*	*Tree*
FF Three Factor	-0.001	-0.001	-0.001	-0.238
FF Five Factor	-0.004	-0.001	-0.004	**0.004**
Carhart Four Factor	-0.005	-0.001	-0.005	-0.005
Ang Volatility	-0.004	-0.001	-0.004	-0.004
Combined	-0.147	-0.001	-0.143	-0.037

Table 10: Prediction Performance of Nonlinear Specifications

	OLS	*LASSO*	*Ridge*	*Tree*
FF Three Factor	-0.006	-0.001	-0.006	-0.447
FF Five Factor	-55.637	-28.475	-55.629	**0.004**
Carhart Four Factor	-0.018	-0.001	-0.01	-0.005
Ang Volatility	-0.032	-0.001	-0.034	-0.004
Combined	-34.473	-73.019	-252.521	-0.037

Finally, the inclusion of more predictors that cover the same equity factors would likely improve the models' ability to predict security returns, for example Gu, Kelly and Xiu (2018) use over 90 predictors. However, data covering the UK is not of the same depth as the US, and although a richer set of predictors could be used, it would not be possible to create a dataset on par with that of American research.

7. Conclusion

This thesis has investigated the extent to which nonlinear statistical methods add value to an equity factor model in the UK. Results are compared mainly in terms of contemporaneous return decomposition models, with an additional investigation into return predictability. The inclusion of nonlinear predictors does not appear to improve model performance, however, the use of shrinkage methods was found to be more performant than a standard OLS model when applied to the linear predictor sets commonly found in the literature. In addition to this, the nonparametric regression tree model performs well in the problem of out-of-sample prediction, supporting current research that finds results along the same lines. Furthermore, as a byproduct of the investigation into nonlinearity, a review is conducted over the persistence of standard literature on equity risk premia by applying similar research methods to a more modern dataset in a non-US market. In particular, the relationships that security returns appear to have with both the size factor and the momentum factor appear to be in contrast to standard theory.

As a result, further research is suggested to complement the work done in this thesis. Extensions could be made to take a deeper look into the value and momentum factors over this time period, as well as to include more factors such as the illiquidity premium, or dividend income. In terms of methodology, factor returns could be value weighted to more accurately assess the return of each factor. Tests of robustness could be performed on the predictions made in section 6 by weighing the economic significance of predictions against practical issues such as trading costs and market impact. Lastly, these predictions could be tested over different time horizons of three or six months where financial noise could be lesser.

As a corollary to the results of this work, certain limiting factors in the research are also acknowledged, Primarily the depth and quality of the data available in the UK market. This has certain anomalous characteristics and potential unreliability in places. This lack of reliability in the data cannot be overcome, and the researcher is forced to trust these anomalies are not too large.

Bibliography

Ang, A., Hodrick, R., Xing, Y. and Zhang, X., 2006. The Cross-Section of Volatility and Expected Returns. *The Journal of Finance*, 61(1), pp.259-299.

Anon, 2020. *What is Low Volatility and Why Does It Matter? - Invesco*. [online] Invesco.com. Available from: https://www.invesco.com/us2/factor-investing/low-volatility-factor-101/ [Accessed 18 Mar. 2020].

Anon, 2020. *Size factor definition - Risk.net*. [online] Risk.net. Available from: https://www.risk.net/definition/size-factor [Accessed 18 Mar. 2020].

Bikhchandani, S. and Sharma, S., 2000. Herd Behavior in Financial Markets: A Review. SSRN Electronic Journal.

Carhart, M., 1997. On Persistence in Mutual Fund Performance. *The Journal of Finance*, 52(1), pp.57-82.

Carlson, B., 2020. *The Momentum Factor: Why Momentum Investing Works*. [online] See It Market. Available from: https://www.seeitmarket.com/the-momentum-factor-why-momentum-investing-works-14544/ [Accessed 18 Mar. 2020].

Fama, E., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance, 25(2), p.383.

Fama, E. and French, K., 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), pp.3-56.

Fama, E. and French, K., 2015. Dissecting Anomalies with a Five-Factor Model. *Review of Financial Studies*, 29(1), pp.69-103.

Gu, S., Kelly, B. and Xiu, D., 2018. Empirical Asset Pricing Via Machine Learning. *SSRN Electronic Journal*.

James, G., Witten, D., Hastie, T. and Tibshirani, R., 2013. *An Introduction to Statistical Learning*. p.222.

Jegadeesh, N., 1990. Evidence of Predictable Behavior of Security Returns. The Journal of Finance, 45(3), pp.881-898.

Jegadeesh, N. and Titman, S., 1993. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance*, 48(1), pp.65-91.

Lintner, J., 1965. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics*, 47(1), p.13.

Miller, M. and Modigliani, F., 1961. Dividend Policy, Growth, and the Valuation of Shares. *The Journal of Business*, 34(4), p.411.

Pioch, A., 2018. *How market frictions live in harmony with persistent factor premia*. [online] Macromatters.lgim.com. Available from: https://macromatters.lgim.com/categories/portfolio-thinking/how-market-frictions-live-in-harmony-with-persistent-factor-premia [Accessed 25 Feb. 2020].

Sharpe, W., 1964. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance*, 19(3), p.425.

Tashman, L., 2000. Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, 16(4), pp.437-450.

Verbiest, E., 2011. Stock Return Prediction and Anomaly Detection by Regression Trees. *SSRN Electronic Journal*.

Appendices

Appendix A: Variable Definitions

The following variables are sourced directly from Thomson Reuters DataStream.

ret	1-month price return in excess of risk-free rate
mv	Market value
allshare	Monthly price return of FTSE All Share
rf	The annualised interest rate on 3-month UK government bonds.
beta	5 Year rolling beta of a stock
so	Total number of shares outstanding
dy	Dividend yield
fcf	Free cash flow yield
opmarg	Operating profit margin
roe	Return on equity
roic	Return on invested capital
debtpct	Total debt as a percentage of book value

The following variables are calculated using the data from DataStream.

bvtmv	Book value to market value ratio
ret_3m	3 month rolling price return
ret_6m	6 month rolling price return
ret_9m	9 month rolling price return
ret_12m	12 month rolling price return
ret_18m	18 month rolling price return
ret_24m	24 month rolling price return
ret_36m	36 month rolling price return
std_3m	3 month standard deviation of return
std_6m	6 month standard deviation of return
std_9m	9 month standard deviation of return
std_12m	12 month standard deviation of return
std_18m	18 month standard deviation of return
std_24m	24 month standard deviation of return
std_36m	36 month standard deviation of return
ind_technology	Dummy variable for companies in technology industry
ind_financials	Dummy variable for companies in financials industry
ind_energy	Dummy variable for companies in energy industry
ind_consumerdiscretionary	Dummy variable for companies in consumer discretionary industry
ind_consumerstaples	Dummy variable for companies in consumer staples industry
ind_basicmaterials	Dummy variable for companies in basic materials industry
ind_industrials	Dummy variable for companies in industrials industry
ind_healthcare	Dummy variable for companies in healthcare industry
ind_utilities	Dummy variable for companies in utilities industry
ind_realestate	Dummy variable for companies in real estate industry
ind_telecommunications	Dummy variable for companies in telecommunications industry

Appendix B: Nonlinear Predictor Sets

FF (1993) Three-Factor Model

$$r_{i,t} - r_{f,t} = \alpha + \begin{pmatrix} \beta_1 MKT_t + \beta_2 MKT_t^2 + \beta_3 MKT_t^3 + \\ \beta_4 HML_t + \beta_5 HML_t^2 + \beta_6 HML_t^3 + \\ \beta_7 SMB_t + \beta_8 SMB_t^2 + \beta_9 SMB_t^3 + \\ \beta_{10}(HML_t \times SMB_t) \end{pmatrix} + \mathrm{u_{i,t}} \qquad (x)$$

Carhart (1997) Four-Factor Model

$$r_{i,t} - r_{f,t} = \alpha + \begin{pmatrix} \beta_1 MKT_t + \beta_2 MKT_t^2 + \beta_3 MKT_t^3 + \\ \beta_4 HML_t + \beta_5 HML_t^2 + \beta_6 HML_t^3 + \\ \beta_7 SMB_t + \beta_8 SMB_t^2 + \beta_9 SMB_t^3 + \\ \beta_{10} UMD_t + \beta_{11} UMD_t^2 + \beta_{12} UMD_t^3 + \\ \beta_{13}(HML_t \times SMB_t) + \\ \beta_{14}(HML_t \times UMD_t) \end{pmatrix} + \mathrm{u_{i,t}} \qquad (x)$$

FF (2015) Five-Factor Model

$$r_{i,t} - r_{f,t} = \alpha + \begin{pmatrix} \beta_1 MKT_t + \beta_2 MKT_t^2 + \beta_3 MKT_t^3 + \\ \beta_4 HML_t + \beta_5 HML_t^2 + \beta_6 HML_t^3 + \\ \beta_7 SMB_t + \beta_8 SMB_t^2 + \beta_9 SMB_t^3 + \\ \beta_{10} CMA_t + \beta_{11} CMA_t^2 + \beta_{12} CMA_t^3 + \\ \beta_{13} RMW_t + \beta_{14} RMW_t^2 + \beta_{15} RMW_t^3 + \\ \beta_{16}(HML_t \times SMB_t) + \\ \beta_{17}(CMA_t \times RMW_t) + \\ \beta_{18}(CMA_t \times SMB_t) \end{pmatrix} + \mathrm{u_{i,t}} \qquad (x)$$

Ang (2006) FF Three-Factor Model with Volatility

$$r_{i,t} - r_{f,t} = \alpha + \begin{pmatrix} \beta_1 MKT_t + \beta_2 MKT_t^2 + \beta_3 MKT_t^3 + \\ \beta_4 HML_t + \beta_5 HML_t^2 + \beta_6 HML_t^3 + \\ \beta_7 SMB_t + \beta_8 SMB_t^2 + \beta_9 SMB_t^3 + \\ \beta_{19} VOL_t + \beta_{20} VOL_t^2 + \beta_{21} VOL_t^3 + \\ \beta_{13}(HML_t \times SMB_t) + \\ \beta_{14}(SMB_t \times VOL_t) \end{pmatrix} + \mathrm{u_{i,t}} \qquad (x)$$

Combined Model

$$r_{i,t} - r_{f,t} = \alpha + \begin{pmatrix} \beta_1 MKT_t + \beta_2 MKT_t^2 + \beta_3 MKT_t^3 + \\ \beta_4 HML_t + \beta_5 HML_t^2 + \beta_6 HML_t^3 + \\ \beta_7 SMB_t + \beta_8 SMB_t^2 + \beta_9 SMB_t^3 + \\ \beta_{10} CMA_t + \beta_{11} CMA_t^2 + \beta_{12} CMA_t^3 + \\ \beta_{13} RMW_t + \beta_{14} RMW_t^2 + \beta_{15} RMW_t^3 + \\ \beta_{16} UMD_t + \beta_{17} UMD_t^2 + \beta_{18} UMD_t^3 + \\ \beta_{19} VOL_t + \beta_{20} VOL_t^2 + \beta_{21} VOL_t^3 + \\ \beta_{22}(HML_t \times SMB_t) + \\ \beta_{23}(HML_t \times UMD_t) + \\ \beta_{24}(CMA_t \times SMB_t) + \\ \beta_{25}(CMA_t \times RMW_t) + \\ \beta_{26}(VOL_t \times SMB_t) \end{pmatrix} + \mathrm{u_{i,t}} \qquad (x)$$

Appendix C: Results Tables

Results from contemporaneous regressions of factor returns on security return, using both the linear and nonlinear predictor set. These are not included due to the lack of economic intuition caused by backward looking predictions.

Table 11: K-Fold CV Results - Linear Predictors

	OLS	*Lasso*	*Ridge*	*Tree*
FF3	-0.031	-0.003	-0.03	-0.005
FF5	**0.038**	-0.003	**0.039**	**0.013**
Mom	-0.005	-0.003	-0.005	-0.015
Vol	**0.056**	-0.003	**0.057**	**0.032**
Combined	**0.061**	-0.003	**0.062**	**0.024**

Table 12: K-Fold CV Results - Nonlinear Predictors

	OLS	*Lasso*	*Ridge*	*Tree*
FF3	-0.117	-0.003	-0.025	-0.005
FF5	-0.025	-0.003	**0.042**	**0.017**
Mom	-0.325	-0.003	-0.005	-0.017
Vol	-0.54	-0.003	**0.054**	**0.029**
Combined	-0.042	-0.003	**0.058**	**0.021**

Fitted polynomial OLS model using nonlinear combined predictor set.

Table 13: Polynomial OLS Coefficients

	Coefficient	*P-Value*
c	0.011	0.142
MKT	201.253	0.233
MKT^2	-197.422	0.150
MKT^3	64.537	0.003
SMB	65.349	0.051
SMB^2	-64.870	0.024
SMB^3	20.971	0.000
HML	-14.898	0.119
HML^2	14.830	0.128
HML^3	-3.830	0.116
RMW	-286.845	0.000
RMW^2	290.206	0.000
RMW^3	-91.711	0.000
CMA	223.441	0.001

CMA^2	-208.556	0.001
CMA^3	71.825	0.000
UMD	0.347	0.311
UMD^2	3.895	0.188
UMD^3	-2.475	0.001
VOL	30.130	0.000
VOL^2	-37.177	0.000
VOL^3	12.082	0.000
HML×UMD	-0.802	0.074
HML×SMB	-2.192	0.004
SMB×VOL	6.964	0.000
CMA×RMW	-18.055	0.009
CMA×SMB	-3.783	0.358

Table 13 shows results from separately controlling for volatility and momentum factors in the FF five-factor model. It shows that volatility captures a lot of the same return as the two quality factors, rendering them insignificant. Meanwhile controlling for the momentum factor does not have the same effect.

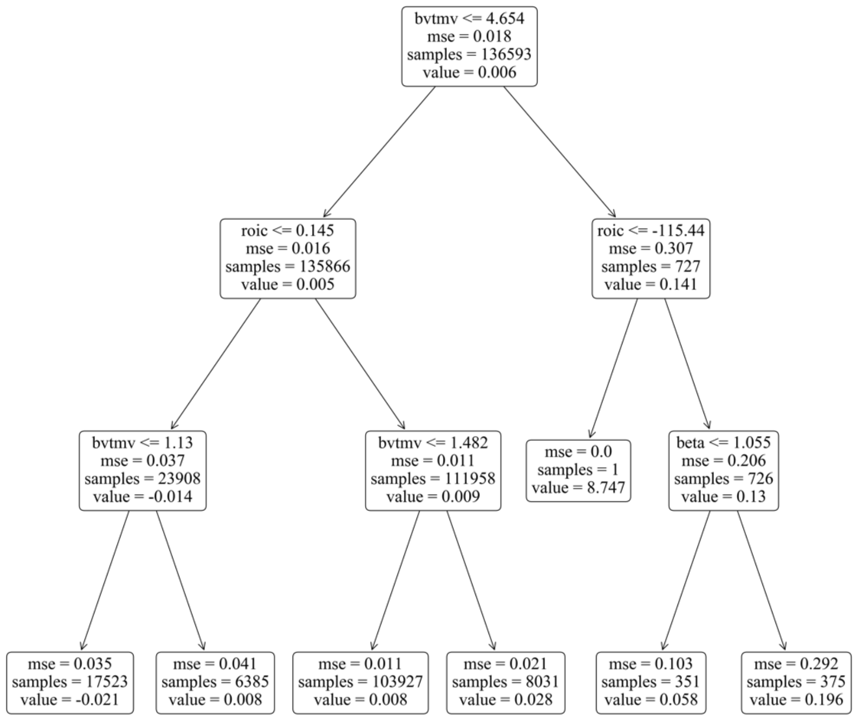
Table 14: FF Five-Factor Model Controlling for Volatility and Momentum

	Coefficient	*P-Value*			*Coefficient*	*P-Value*
c	0.002	0.125		*c*	0.022	0.000
MKT	0.270	0.000		*MKT*	0.275	0.000
SMB	-0.499	0.000		*SMB*	-0.441	0.000
HML	0.474	0.000	vs	*HML*	0.203	0.042
RMW	-0.233	0.151		*RMW*	-0.892	0.000
CMA	0.034	0.064		*CMA*	0.051	0.036
VOL	-0.597	0.000		*UMD*	-0.092	0.010

Appendix D: Prediction Results

Regression tree does not include any of the profitability metrics in prediction.

Figure 18: Regression Tree of FF Five Factor Model for Prediction



Appendix E: Code

All data analysis was performed in Python. Data cleaning was carried out with the *pandas* and *numpy* libraries, models were fitted and tuned using *sklearn*, and graphs were plotted using *matplotlib* unless otherwise stated. All code can be found at www.github.com/BarnabyN/EconomicsDissertation_Code.