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

April 2020

University of Bath

Department of Economics

Acknowledgements

* Patrick
* Andrzej
* Dad

Abstract

what is the goal of the paper?

what does the paper discover/prove/conclude?

**Table of Contents**

[Acknowledgements 1](#_Toc36291318)

[Abstract 1](#_Toc36291319)

[1. Introduction 3](#_Toc36291320)

[2. Literature Review & Theory 5](#_Toc36291321)

[2.1. Efficient Market Hypothesis 5](#_Toc36291322)

[2.2. Equity Risk Premia 6](#_Toc36291323)

[2.2.1. The Market 6](#_Toc36291324)

[2.2.2. Size and Value 7](#_Toc36291325)

[2.2.3. Momentum 7](#_Toc36291326)

[2.3. Quality 8](#_Toc36291327)

[2.3.1. Volatility 9](#_Toc36291328)

[2.4. Non-Linearity 10](#_Toc36291329)

[2.4.1. Generalised Linear Models 10](#_Toc36291330)

[2.4.2. Supervised Machine Learning 10](#_Toc36291331)

[2.4.3. Deep Learning 11](#_Toc36291332)

[3. Data 11](#_Toc36291333)

[3.1. Collection & Sources 11](#_Toc36291334)

[3.2. Constructing Factor Returns 12](#_Toc36291335)

[3.3. Defining Predictor Sets 14](#_Toc36291336)

[3.4. Evidence of Nonlinearity 15](#_Toc36291337)

[4. Methodology 19](#_Toc36291338)

[4.1. Models Estimated 19](#_Toc36291339)

[4.1.1. Linear Regression 19](#_Toc36291340)

[4.1.2. Polynomial Regression 20](#_Toc36291341)

[4.1.3. LASSO Regression 20](#_Toc36291342)

[4.1.4. Ridge Regression 21](#_Toc36291343)

[4.1.5. Regression Tree 21](#_Toc36291344)

[4.2. Model Training 23](#_Toc36291345)

[4.3. Hyperparameter Tuning 25](#_Toc36291346)

[5. Results 27](#_Toc36291347)

[6. Making Predictions 33](#_Toc36291348)

[7. Conclusion 34](#_Toc36291349)

[Bibliography 35](#_Toc36291350)

[Appendix 37](#_Toc36291351)

1. Introduction

For as long as the stock market has existed, its participants have been searching for ways to accurately assess the expected return of a stock. The main motivation is usually to make money, but beyond this an accurate evaluation of a stock’s expected return can be informative for the company’s management in choosing a discount factor for projects it is considering undertaking. Along the same lines, a security’s expected return can be used 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 factor exposures. The generalized linear specification is shown in equation [x].

Where each is the return of factor in period , and is the exposure of stock to factor . For a factor model to be complete, must statistically indifferent from 0, and 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.

this thesis will look at both contemporaneous models and prediction models

some nonlinear models (like regression trees) can be rather intuitive in modelling relationships

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 on in the following section, equity factors have generally been modelled and tested using a linear specification 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. We may wish to use more advanced statistical models to estimate the expected return of a stock, allowing for the relaxation of the linearity constraint, as well as allowing for interaction terms between predictors. As a result of this fact, the goal of this thesis is to investigate the extent to which nonlinear statistical models perform better in decomposing security returns into their factors in the UK, as well as using these factors for return prediction.

Some analysis of this type has been done over the years, with the idea for this thesis being born 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. I believe that there should be an attempt made to do the same on UK data, allowing for a comparison of non-linear multi factor models across the two countries. That being the case, UK data has a shorter time scale and less detail, which this thesis acknowledges.

The rest of this paper is broken down into [X] main sections. Firstly, section [y] provides a summary of the research that has been done on equity factors and their nonlinearity. Following this, section [y] elaborates on the models applied in this thesis, as well as concluding the key findings from their application.

1. 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, meanwhile section 2.3 summarises the literature on machine learning and its application to modelling risk premia.

* 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 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.

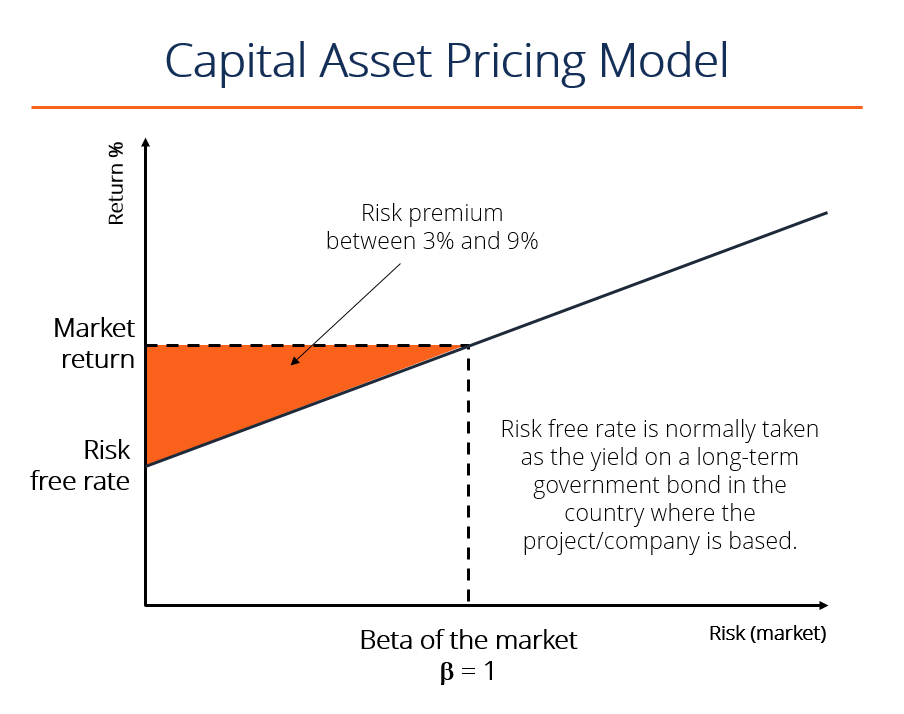
The implications of the EMH for equity factors are as follows; no investor, statistician or trader can use the set of publicly available information to correctly and consistently forecast asset returns.

* 1. Equity Risk Premia
     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.

The beta of a stock represents its riskiness relative to the market and can be defined by equation 2 below, or more commonly by the regression coefficient produced when running OLS on equation 1.

In the risk-reward space, the CAPM model is represented by the Security Market Line (SML), depicted in figure [y]. This line is upward sloping, indicating that taking on additional risk entitles an investor to a higher expected return. By taking the risk of the market as defined by a beta of 1, figure [y] 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.



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.

* + 1. 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 (SMB) factor and value (HML) factor. They 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 [x].

Firstly, *the Size factor* is represented by the difference in average return on the smallest and largest 20% 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 she 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, there is no doubt that the size factor is a result of investors being compensated for taking additional risk.

Secondly, *the Value factor* is calculated as the difference in average return on the 20% of 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.

* + 1. Momentum

Momentum is the observation that stocks with strong 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 at short lags such 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. 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) followed up on the paper, evaluating 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. The model 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.

talk more about carhart model

* + 1. Quality

Many years later, Fama and French (2015) added to their three-factor model by incorporating a Quality factor exposure. Quality implores managers and investors to buy good companies who invest well and make solid profits.

Firstly, *low investment* as defined by the CMA factor equals the return of companies that invest conservatively minus the return of companies that invest aggressively. 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 functional here, and hence why any excess return is not bought up immediately.

The two components of the quality factor 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.

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

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

Robust profitability refers a high value of in equation [x], while conservative investment implies a low . Both lead to a higher value of in equation [x] and in equation [x].

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

* + 1. Volatility

Finally, Andrew Ang (2006) showed that companies have an exposure to market volatility, and those with a lower sensitivity would have higher expected returns. Specifically, he used the VIX index, a measure of market volatility, to show that companies with less exposure to changes in the index saw statistically significant outperformance in future periods.

Low volatility is a rather confusing factor in that it appears to directly contrast the risk-reward framework at the centre of financial markets. The main theory behind its existence is what is known as the lottery effect (What is Low Volatility and Why Does It Matter? - Invesco, 2020). The idea is that in order to achieve a higher expected return, investors go looking for high risk companies that can offer these higher returns. As a result, lower risk companies are underbought and hence offer a return premium to investors as a result of being avoided in the first place.

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 [x] and [x].

* 1. Non-Linearity

So far all of the models put forward use a linear specification 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.

explain why there is less literature on this section (if it works why would you publish it)

* + 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 [x].

As a result, these models allow for nonlinearities between each and . 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.

* + 1. Supervised Machine Learning

what is supervised machine learning and how does it compare to unsipervised

Building on generalised linear models is common in machine learning literature. Not much of the literature around equity factors has implemented machine learning models, however, those that did 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.

* + 1. 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), meanwhile 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. 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. The reader is encouraged to survey the aforementioned literature for an idea of the state of research in this area.

1. 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 3.1. Following this, 3.3 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.

* 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 at this stage of the process 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 table [X] of the appendix.

Companies are deemed eligible for inclusion in regressions in period if the independent variables used in the regression are available in period , and return of the security (the dependent variable) available in period . 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.

* 1. Constructing Factor Returns

Before testing various models, factor returns must be constructed in the same fashion as the papers from which they originate. The factors used in this thesis encompass all those included in the section 2.2. Table [x] details how each factor is calculated.

Table 1 - Definitions of Factor Returns

|  |  |
| --- | --- |
| *MKT* |  |
| *HML* |  |
| *SMB* | Return of 20% of companies with smallest market value – Return of 20% of companies with largest market value |
| *CMA* | Return of 20% of companies with highest return on invested capital – Return of 20% of companies with lowest return on invested capital |
| *RMW* | Return of 20% of companies with highest operating profit margin – Return of 20% of companies with lowest operating profit margin |
| *UMD* | Return of 20% of companies with highest return over last twelve months – Return of 20% of companies with lowest return over last twelve months |
| *VOL* | Return of 20% of companies with lowest return volatility over last twenty-four months – Return of 20% of companies with highest return volatility over last twenty-four months |

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 [y] shows the returns of each factor relative to the market since 2010, while table [x] specifies the risk adjusted performance of each factor.

Figure 1 - Factor Performance Relative to Market

A close up of a mans face

Description automatically generated

|  |  |
| --- | --- |
|  | *Sharpe Ratio* |
| *MKT* | 0.206868 |
| *HML* | 0.345043 |
| *SMB* | 0.107997 |
| *CMA* | 0.757136 |
| *RMW* | 0.409020 |
| *UMD* | 0.312279 |
| *VOL* | 0.016650 |

comment on factor performance

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%.

Then the higher order polynomials are calculated as follows.

* 1. Defining Predictor Sets

The linear model uses a linear predictor set identical to those in the pre-defined models, while the nonlinear models use a predictor set in which each variable in the linear predictor set is included up to the cubic polynomial. The linear form of each model is specified according to equations [x] to [x] respectively, with the addition of nonlinearities where appropriate. A full list of the nonlinear predictor sets can be found in appendix [x].

*FF (1993) Three-Factor Model*

*Carhart (1997) Four-Factor Model*

*FF (2015) Five-Factor Model*

*Ang (2006) FF Three-Factor Model with Volatility*

Finally, in order to compare the impact that each factor has on the coefficients of other factors, a combined model is constructed consisting of the union of all the above models.

*Combined Model*

* 1. Evidence of Nonlinearity

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. The figures below show the relationship between various predictors of asset return and the asset return itself. In most 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.

make these graphs prettier (larger text)

A picture containing photo, white, covered, table

Description automatically generated

A picture containing text, photo, white, light

Description automatically generated

A picture containing text, photo, white, computer

Description automatically generated

A picture containing text, photo, computer, white

Description automatically generated

explain the cubic curve shape and the intuition behind why it may be a better fit (eg. too good value usually has a reason for it, extremely bad value can be forecasting future growth)

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. Figure [x] shows a polynomial regression of the FF five factor model on returns over the first 80% of the dataset.

A close up of text on a white background

Description automatically generated

Finally, a correlation matrix over the set of constructed factors suggest that there may be interaction effects between them. The main results from these matrices can be rationalised from what is known about equity 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 2 - Factor Returns Correlation Matrix



1. Methodology

This section will discuss the models that are used, as well as how each is fitted and validated. Section 4.1 outlines the different models that are to be compared, while sections 4.2 and 4.3 outline the method via which these models are fitted, and their tuning parameters chosen.

elaborate on this: what models are tested, how are they tested, what are the test results

The models summarised in section 2.2 are used as the foundation upon which this thesis builds.

* 1. Models Estimated

The goal of this paper is to evaluate if models that allow for nonlinearity in the relationship between common equity factor returns and security returns improve upon the standard linear regression. As a result, a variety of models are considered, and their specifications summarised in this section.

* + 1. Linear Regression

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 [x].

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

* + 1. Polynomial Regression

The first nonlinear model implemented is the polynomial regression, which builds on the linear regression by allowing for quadratic and cubic polynomials of variables to be used as predictors.

Firstly, a cubic polynomial predictor set is created from the existing data simply by raising all variables up to the third power. Each model specified above is expanded to include all predictors up to a cubic power, hence permitting a nonlinear relationship.

This model uses the various nonlinear predictor sets defined in section 3.3.

No strategic process was used to choose a cubic polynomial, and it is chosen largely due to statistical conventions when running polynomial regressions. However, the argument could be made that a cubic polynomial regression is 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.

The goal of the polynomial regression is the same as that of the linear regression, namely, to minimise the sum of squared residuals as defined in equation [x] of section 4.1.1. Hence outliers are just as important in this scenario, however the idea is that by permitting a nonlinear relationship between predictors and returns the model will be a better fit.

* + 1. 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 [x].

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

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.

A picture containing umbrella

Description automatically generated

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

* + 1. 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 [x].

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 [x]. 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.



* + 1. 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 [x] of section 4.1. The same splitting approach is the applied to each of the subgroups that the data has been split into. 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 1, 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 1.

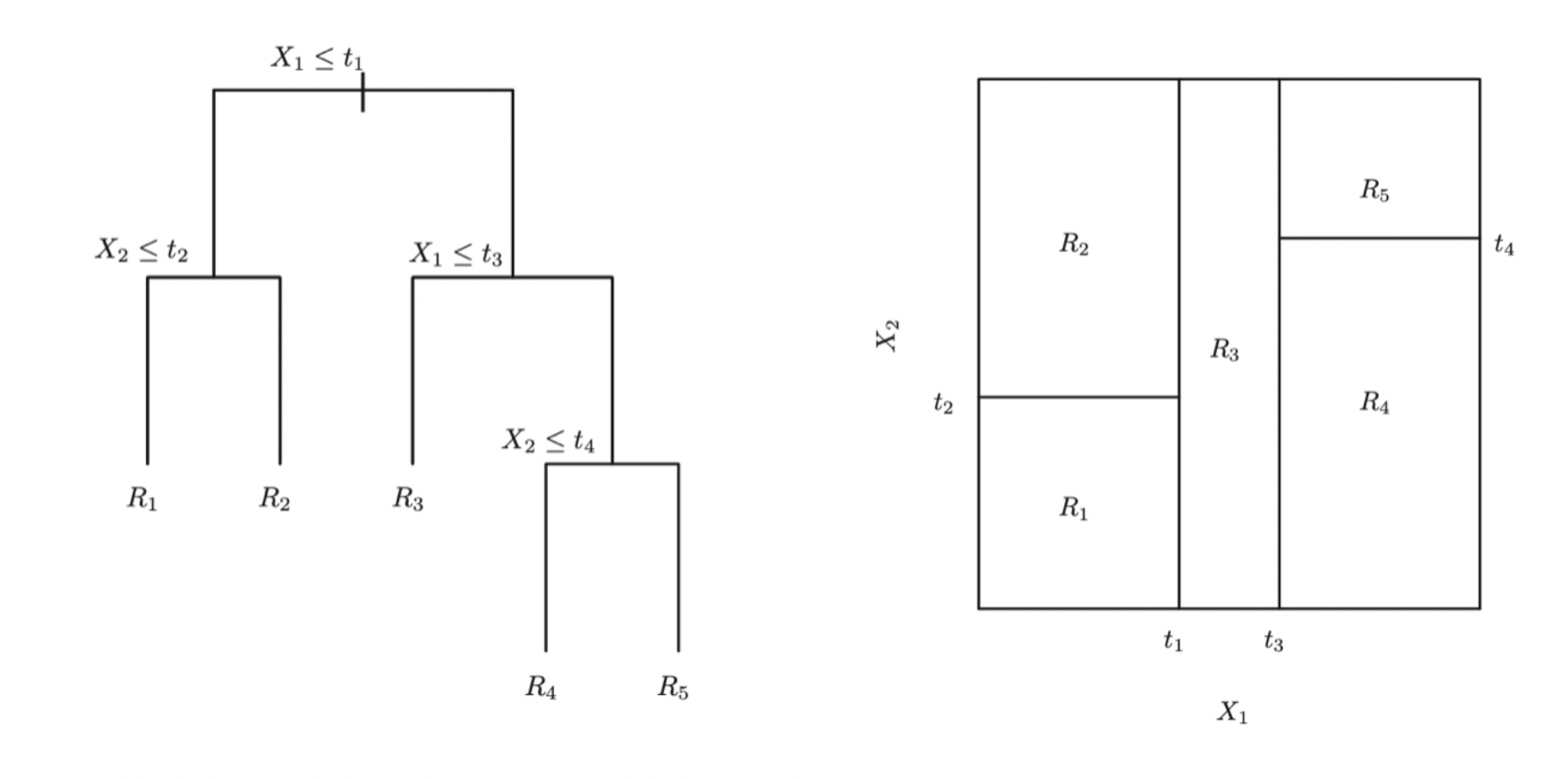
Figure 3 - Regression Tree Output of Carhart Four Factor Model

A screenshot of a cell phone

Description automatically generated

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 explanation behind the parameter choice made in this thesis is 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.



* 1. Model Training

Traditionally, a process of k-fold cross validation is used to train the model and evaluate its true performance. K-fold cross validation involves dividing the dataset into segments, or “folds”. One by one each fold is temporarily removed from the dataset. The model is then trained on the other 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 [x] and is a common error measure in both statistics and machine learning.

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

Below is a graph showing the performance as measured by MSE in modelling the FF five factor relationship over different numbers of folds.

Figure 4 - Plot of average MSE vs number of folds used in k-fold CV

A picture containing table, room, white, standing

Description automatically generated

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 in not only lacks intuition but it may also generate distorted estimates of out of sample performance.

Consequently, in addition to k-fold cross validation, a process of rolling origin (or “walk forward”) evaluation (Tashman, 2000) is also used. This involves training the model on an initial subset of the data and testing it on the proceeding fold. This process is then repeated sequentially so as to maintain the temporal order of the data, ensuring that all model predictions are forward looking in time. Figure 2 shows the performance of each model

Figure 5 - Plot of average MSE vs number of folds in rolling origin CV

A close up of a map

Description automatically generated

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 at larger numbers of folds in Figure 3.

In order to strike a balance of performance between k-fold cross validation and rolling origin evaluation, five folds are used for both.

* 1. Tuning Parameter Optimisation

There are three tuning parameters that the models must be optimised with respect to. The first two are the 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.

Figure 6 - Ridge Regression Hyperparameter Tuning

A picture containing indoor, table, white, computer

Description automatically generated

Figure 7 - Regression Tree Parameter Tuning

A picture containing text, table, computer, room

Description automatically generated

1. Results

When evaluating each model, is used as a measure of performance. is defined as the percentage variation in the dependent variable that is explained by the model and is specified by equation 5.1.

The results from training the models using k-fold cross validation method are shown below in table 2. As expected, values are low due to the high degree of noise present in financial market data.

Table 2 – Average Performance from K-Fold Cross Validation

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Linear* | *Polynomial* | *Lasso* | *Ridge* | *Tree* |
| *FF3\_* | -0.031 | -0.117 | -0.003 | -0.025 | -0.005 |
| *FF5\_* | **0.038** | -0.025 | -0.003 | **0.042** | **0.017** |
| *mom\_* | -0.005 | -0.325 | -0.003 | -0.005 | -0.017 |
| *vol\_* | **0.056** | -0.54 | -0.003 | **0.054** | **0.029** |
| *combined\_* | **0.061** | -0.042 | -0.003 | **0.058** | **0.021** |

Table 3 – Average from Rolling Origin Evaluation

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Linear* | *Polynomial* | *Lasso* | *Ridge* | *Tree* |
| *FF3* | -0.014 | -0.042 | -0.001 | -0.026 | -0.026 |
| *FF5* | **0.033** | -0.658 | -0.001 | **0.024** | **0.008** |
| *Mom* | **0.005** | -0.598 | -0.001 | **0.009** | **0.006** |
| *Vol* | **0.056** | -0.333 | -0.001 | **0.057** | **0.017** |
| *Combined* | **0.06** | -0.426 | -0.001 | **0.057** | **0.003** |

comparison of linear to non-linear

Tables 2 and 3 allow us to perform our first comparison of nonlinear models to the linear model. The performance across different predictor sets does not initially appear to show any improvement of fit.

Linear models are simpler and easier to interpret, but the regression coefficients on the non-linear models suggest that there is potential for nonlinear coefficients to be included in the model.

discussion of regression coefficients

In addition to measuring model performance using , the extent to which a factor model captures all return variation can be evaluated via a brief investigation of the regression intercept and residuals, primarily the former. FF (1993) showed that the estimated intercept in their three-factor model was statistically insignificant from 0. In this thesis, tables 4 and 5, which show the regression coefficients of the OLS model using both the linear and nonlinear combined predictor sets, both have intercepts that are statistically insignificant from zero, suggesting that the models are potential candidates for complete factor models in the UK.

Table 4 - Linear Regression Coefficients

|  |  |  |
| --- | --- | --- |
|  | *Coefficient* | *P-Value* |
| *c* | 0.003 | 0.108 |
| *MKT* | 0.262 | 0.000 |
| *SMB* | -0.453 | 0.000 |
| *HML* | 0.336 | 0.000 |
| *RMW* | -0.240 | 0.110 |
| *CMA* | 0.175 | 0.176 |
| *UMD* | -0.109 | 0.000 |
| *VOL* | -0.595 | 0.000 |

Table 5 - Polynomial Regression 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 |
| *HMLUMD* | -0.802 | 0.074 |
| *HMLSMB* | -2.192 | 0.004 |
| *SMBVOL* | 6.964 | 0.000 |
| *CMARMW* | -18.055 | 0.009 |
| *CMASMB* | -3.783 | 0.358 |

lasso and ridge coefficients

Continuing the investigation into the regression coefficients, the coefficients of both the LASSO and ridge are shown in table 6. Remarkably, the LASSO regression reduces all coefficients to zero, predicting a constant monthly return that is the same across all securities and time periods. This prediction of 0.7% monthly return amounts to an expected return of 8.7% per annum.

Table 6 - LASSO and Ridge Regression Coefficients

|  |  |  |
| --- | --- | --- |
|  | *LASSO* | *Ridge* |
| *c* | 0.007 | 0.006 |
| *MKT* | 0 | 0.353 |
| *MKT 2* | 0 | 0.271 |
| *MKT 3* | 0 | -0.217 |
| *SMB* | 0 | 0.017 |
| *SMB 2* | 0 | -0.027 |
| *SMB 3* | 0 | -0.090 |
| *HML* | 0 | 0.006 |
| *HML 2* | 0 | 0.017 |
| *HML 3* | 0 | 0.114 |
| *RMW* | 0 | 0.041 |
| *RMW 2* | 0 | 0.015 |
| *RMW 3* | 0 | -0.125 |
| *CMA* | 0 | 0.083 |
| *CMA 2* | 0 | 0.095 |
| *CMA 3* | 0 | -0.039 |
| *UMD* | 0 | 0.053 |
| *UMD 2* | 0 | 0.087 |
| *UMD 3* | 0 | -0.121 |
| *VOL* | 0 | -0.089 |
| *VOL 2* | 0 | -0.064 |
| *VOL 3* | 0 | -0.084 |
| *HMLUMD* | 0 | 0.047 |
| *HMLSMB* | 0 | -0.073 |
| *SMBVOL* | 0 | -0.164 |
| *CMARMW* | 0 | 0.003 |
| *CMASMB* | 0 | 0.098 |

Meanwhile, the ridge regression coefficients are shown in the right-hand column of table 6, comparing these to the standard polynomial regression with no shrinkage we can see which variables are of most or least importance. Figure [x] shows the difference in regression coefficients between the polynomial and ridge regressions. Positive bars indicate that the ridge regression values the predictor more highly than standard OLS, and a negative bar indicates the opposite.

Figure 8 - Coefficient Differences Between Ridge and OLS

A screenshot of a computer

Description automatically generated

show how coefficients change over different tuning parameter values?

A close up of a map

Description automatically generated

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, in the prediction models of section 6 these models lose a significant amount of value 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. This can be overcome however, by standardizing the predictors.

regression tree feature importance

Finally, an analysis of the coefficients of the regression tree model is conducted. When fitting a regression tree, each predictor has a certain degree of importance. Put simply this importance is defined as the sensitivity of the overall prediction to changes in this variable.

Figure 9 shows the predictor importance coefficients of the nonlinear combined model.

Figure 9 - Feature Importances in Nonlinear Combined Model

A close up of a logo

Description automatically generated

Figure 10 - Regression Tree of Nonlinear FF 5-Factor Model

A screenshot of a cell phone

Description automatically generated

Overall analysis of the respective importance of each variable in each model.

A close up of a map

Description automatically generated

limitations of results

To conclude this section, 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 is where most research on risk premia lies. 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 in the first place.

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.

1. 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 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. Therefore, the underlying variables used to construct factor returns are regressed on instead. For example, instead of the HML factor, the underlying book-to-market value is regressed instead.

Regression trees appear to be particularly effective in return prediction as opposed to return decomposition. These predictions also appear to be robust over different prediction horizons, with table [x] showing the results of predicting 3-month future returns.

Table 7 - Performance of 1-Month Predictions

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Linear* | *Polynomial* | *Lasso* | *Ridge* | *Tree* |
| *FF3* | -0.001 | -0 | -0.002 | -0 | -0.005 |
| *FF5* | -0.004 | -0.246 | -0.337 | -0.246 | **0.011** |
| *mom* | -0.001 | -2.932 | -0.002 | -5.445 | -0.033 |
| *vol* | -0.003 | -0.018 | -0.002 | -0.02 | -0.006 |
| *combined* | -0.115 | -0.325 | -0.482 | -0.816 | -0.001 |

Table 8 - Performance of 3-Month Predictions

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Linear* | *Polynomial* | *Lasso* | *Ridge* | *Tree* |
| *FF3* |  |  |  |  |  |
| *FF5* |  |  |  |  |  |
| *Mom* |  |  |  |  |  |
| *Vol* |  |  |  |  |  |
| *Combined* |  |  |  |  |  |

1. Conclusion

could extend to more data and factors: illiquidity, value weight factors, try more ml models (neural networks)

how significant are these results in the field of equity risk premia

nobody has done this before, so it does show to an extent how the data in the uk can be used

also provides the basis for a uk equity factor model, of which there are also fewer

you may lose more in interpretability than you gain in predictability

Overall, 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.

Bibliography

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

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.

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.

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

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

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

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.

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

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].

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.

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

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

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

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].

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

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].

Appendix

**Python Code**

The code used to run the regression models and machine learning models. Programmed in Jupyter notebook.

**Variables**

The following variables are sourced directly from Thomson Reuters DataStream.

|  |  |
| --- | --- |
| **Name** | **Description** |
| *ret* | 1-month price return in excess of risk free rate |
| *mv* | Market value |
| *allshare* | Monthly price return of FTSE All Share |

The following variables were calculated using the data in the above table.

|  |  |
| --- | --- |
| **Name** | **Description** |
| *ret\_12m* | 3-month price return in excess of risk free rate (gilt3m) |
| *gilt3m* | The annualised interest rate on 3-month UK government bonds. Used as risk free rate |
|  |  |

**Nonlinear Predictor Sets**

|  |  |
| --- | --- |
| *Model* | *Predictors* |
| FF3 |  |
| FF5 |  |
| Carhart 4 |  |
| Ang Vol |  |

*FF (1993) Three-Factor Model*

*Carhart (1997) Four-Factor Model*

*FF (2015) Five-Factor Model*

*Ang (2006) FF Three-Factor Model with Volatility*