An Investigation into the Performance of Machine Learning Models in Pricing Equity Risk Premia in the UK

ES30029 - Final Year Research Project

Barnaby Napier, 14041

2020

Contents

[Abstract 2](#_Toc31614514)

[Introduction & Background 2](#_Toc31614515)

[Literature Review & Theory 2](#_Toc31614516)

[Data 3](#_Toc31614517)

[Methodology & Models 4](#_Toc31614518)

[Results 4](#_Toc31614519)

[Results 4](#_Toc31614520)

[Limitations 4](#_Toc31614521)

[Conclusion 4](#_Toc31614522)

[Bibliography 4](#_Toc31614523)

[Appendix 5](#_Toc31614524)

# Acknowledgements

# Abstract

# Introduction & Background

* Equity factors and their uses in markets and finance
* Why they exist
* What use a good model of equity factors is
* Why we may want to use modern techniques to investigate these relationships
  + The linear relationship assumption leads to an implicit restriction not only of no quadratic relationship but also of no interaction terms
* ML has been applied to the USA where the data is more abundant and cleaner
* The UK has both a shorter history of data and less detailed data (this must be kept in consideration)
* Hence many researchers keep clear of the UK market
* I acknowledge these limitations, yet believe there should still be some published attempt made at modelling UK equity factors using more modern statistical techniques which relax some of the assumptions of linear regressions

# Literature Review & Theory

## Efficient Market Hypothesis

* Fama and Malkiel
* Information discovery is random and hence the process of a stock price also follows a random process
* This said that all available information about a company is immediately incorporated into its stock price
* Hence there is no way to outperform the market consistently (would just be due to luck)
* This is supported by the lack of persistence in fund manager performance

## Factor Models

### CAPM

* Henry Sharpe suggested that stocks are priced according to their level of risk, as measured by their variation relative to the market
* This was termed “market beta” and is used widely throughout the industry today as a measure of systemic risk in a stock
* Beta would be the beta (regression coefficient) if you ran the above regression

### Fama French 3 Factor Model

* Eugene Fama and Kenneth French built on the work of Sharpe, concluding that the market was one of multiple “factors” influencing equity returns
* During the late 1900’s they researched and built various models to explain stock returns using these “style factors”
* Fama-French 3 factor model: Market, Size, Value

### Fama-French 5 factor model

* Two factors modelling quality
  + + Low investment
  + + Profitability
* These equity factors have their roots in financial economic theory, and are backed up by traditional asset pricing models
* The illusive term “alpha”, which defines a portfolio managers return above that of some benchmark was being worn away
* Fama and French had shown that the much of a fund manager’s performance can be explained through their models

## Market Anomalies

### Momentum

* As alpha became harder to find, managers were on the lookout for true anomalies which could not be explained by factors such as value and size.
* Momentum is by far the most famous. The idea that high past returns will lead to high future returns and vise versa (12 month time scale)
* Momentum implies that market returns have some memory (i.e there is autocorrelation in stock returns)

### Volatility

* Other anomalies include Seasonality and Volatility
* The difference between an anomaly and an equity factor is a little fuzzy however

## Non-Linearity

### Non-Linear Regression Models

* Polynomial regression

### Machine Learning

* The advance of machine learning since around 2010 has been accompanied by a drastic increase in its application to finance and the markets
* Machine learning models are broadly broken down into classification and regression
* Not much of the literature has used machine learning in relation to equity factors
* The few papers which have used it have found promising results
* Main models that I have found being used in relation to equity factors
  + Neural networks
  + Random forests & Regression trees
  + Partial Least Squares
  + Principal Component Regression
  + Generalised linear regression
  + Penalised linear regression

# Analysis

## Data

### Collection & Sources

This section describes the data used in subsequent analysis. It consists of monthly stock level data from all companies in the FSTE All-Share, beginning on 31/12/1995, and is sourced from Thomson Reuters DataStream. Monthly sampling was chosen due to

* Do some quality tests on the data?
  + Stationarity
  + Normality
  + Autocorrelation
  + Etc

**Appendix A**

|  |  |
| --- | --- |
| *ret* | 1 month price return |
| *ret\_3m* | 3 month price return |
| *…* | … |

### Exploratory Analysis

A brief investigation into the data suggests that statistical models which allow for non-linear relationships between independent and dependent variables may be better suited to the problem of asset pricing. The below scatter charts show the relationship between various predictors of asset return and the asset return itself. In most cases the argument can easily be made that a linear model would fail to effectively model the respective relationship.

A close up of a map

Description automatically generated

A close up of a map

Description automatically generated

A screenshot of a cell phone

Description automatically generated

* Regressions including polynomial terms

Secondly, correlation matrices over different time frames 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: Small cap has more momentum …

A screenshot of a cell phone

Description automatically generated

## Models & Methodology

### Linear Models

* Linear regression specification and performance

### Non-Linear Models

### Machine Learning Models

* ML models tested
  + Which ones performed well?
  + Why do you think they performed well?
  + Stats behind the models
* Approach bettering the model
  + Hyperparameter tuning
  + Regularization

## Results

### Results

* R-squared (expect values to be low as this is financial data)
* Mean squared error (MSE) – Used in Patrick’s lectures so he will like it being in here

### Limitations

* Data availability
* Economic significance (investors most likely cannot use this to their advantage)

# Conclusion

# Bibliography

# Appendix

* Code