The Predictive Nature of Short Interest on Market Returns and the Effect of Short Selling on Volatility, Liquidity and Price Discovery with Some International Evidence.

HARIHAR VIRENDRA PATEL

A thesis submitted in partial fulfilment of the requirements of the University of Greenwich for the Degree of Doctor of Philosophy

APRIL 2020

DECLARATION

I certify that the work contained in this thesis, or any part of it, has not been accepted in substance for any previous degree awarded to me, and is not concurrently being submitted for any degree other than that of Doctor of Philosophy being studied at the University of Greenwich. I also declare that this work is the result of my own investigations, except where otherwise identified by references and that the contents are not the outcome of any form of research misconduct.

Harihar Virendra Patel

Dr Francesco Guidi (first supervisor)

Prof Aleksandar Stojanovic (second supervisor)

ACKNOWLEDGEMENTS

I would like to thank my first supervisor Dr Francesco Guidi for his insightful comments and help in the production of this thesis. I would also like to thank second supervisor Prof Aleksandar Stojanovic and Dr V L Raju Chinthalapati for their additional advisory comments throughout the years in regards to this thesis. Without their expertise and knowledge in the field of finance and econometrics, this thesis would not have been possible. I would also like to thank Dr Larry Su for helping me refine my research proposal to a standard in which I was able to obtain the Vice Chancellor (VC) Scholarship at the University of Greenwich in regard to this project.

I would also like to thank Dr Guido Conaldi, Dr Sven Kuenzel, Dr Anne-Marie Coles and Prof Alan Reed for the research methods training they provided during the preliminary stages of this PhD. The research training given in quantitative methods and qualitative methods has been invaluable. I would also like to thank Dr Emily Salines, Dr Timos Almpanis, Dr Martin Compton and Dr Danielle Tran for the assessment, feedback and teaching training they provided. I would like to also thank Dr Gabriella Cagliesi and Dr Sohan Sarwar for their feedback during the transfer viva, their comments have been invaluable in shaping the structure and direction of this thesis. I would also like to thank fellow PhD candidate in the Department of International Business and Economics, Timo Van der Linden for his help during the transfer viva and further comments regarding this PhD. I would also like to thank Mrs Gillian Haxell for the great deal of internal administrative duties she has helped me with regarding administrative forms, the transfer viva and final viva. She has been outstanding when it comes to administration in the business school. I would also like to thank Prof Bruce Cronin, the Director of Research at the University of Greenwich Business School, for choosing me above many other able and willing candidates to produce research for this PhD and deliver this thesis, I hope the findings are of great interest and value and I sincerely wish the knowledge this PhD delivers inspires others to develop research into this field further.

I would also like to thank external examiner Prof Daniel Ladley of the Univesity of Leicester and internal examiner Dr Giuseppina Madonia for their comments during the final viva and taking time in reading my thesis. Their extensive knowledge in finance and econometrics is invaluable in refining this thesis in regards to minor amendments that have had to be made for completion of this thesis to the required academic standard.

I would also like to thank the staff in the University of Greenwich Library and University of Greenwich IT Department who have been very helpful in matters in regard to software, datasets and textbook loans. In particular I would like to thank Mrs Rosalind Smith for her help in regards to Thomson Reuters Eikon and DataStream software, which has been crucial to the data I have been using across all research questions. Finally, I would like to thank all my friends and family who have been greatly supportive during this PhD.

ABSTRACT

I look to explore the findings of Boehmer et al. (2010) and in particular test across model specification investment horizon and across countries. This in turn means I look at whether short sellers are informed traders and if there is evidence that short sellers engage in market manipulation. I also look at whether and to what extent short sales affect liquidity, price discovery, volatility and cross-section of stock returns. The main novelties of research of this PhD are that I employ a new more efficient asset pricing model to the findings of Boehmer et al. (2010), I explore also a time period after publication to observe market efficiency and that I explore an international market. I am also one of few studies to study volatility, liquidity and price discovery in regards to a short sale ban in the UK.

Firstly, I look at whether the strategy of Boehmer et al. (2010) is valid when another most recent and efficient model is used to adjust for risk premium. Boehmer et al. (2010) used the Fama and French Three Factor Model with Momentum to adjust for risk premium, in my case I use the Fama and French Five Factor model to adjust for risk premium. The strategy of Boehmer et al. (2010) involves going long the top percentile of stocks ranked by short interest and going short the bottom percentile of stocks ranked by short interest and rebalancing each month based on new short interest data. I find that heavily shorted stocks underperform lightly shorted stocks. The Fama and French Five Factor model holds a high positive alpha for lightly shorted portfolios on top of both higher excess and raw returns for lightly shorted portfolios compared to heavily shorted portfolios. However, the short component of the Boehmer et al. (2010) strategy yields a positive return, therefore I advise to go long the top five percentile of stocks. The Fama and French Five Factor model does indicate there is good news in short interest albeit without a short component.

Secondly, I look at whether the strategy of Boehmer et al. (2010) is still valid after publication in the US stock market or whether an efficient market has caused arbitrage to take place and make this strategy redundant. I again find that heavily shorted stocks underperform lightly shorted stocks, however the strategy of Boehmer et al. (2010) is not completely valid as the short component of the strategy yields a positive monthly return. It is still however valid to go long the top five percentile of stocks and I believe that this underperformance of heavily shorted stocks means arbitrage has not taken place.

Thirdly, I also look at whether the strategy of Boehmer et al. (2010) is valid in another OECD country such as Canada. This is used as an indicator of the international OECD market based on short interest data availability. I find again that heavily shorted stocks underperform lightly shorted stocks, in line with findings consistent with previous literature. However, the strategy of Boehmer et al. (2010) is again completely not valid as the bottom five percentile of stocks have a positive monthly return. I advise going long the top five percentile portfolio as the best strategy, in line with my findings regarding arbitrage in the US stock market. There is no valid reason in holding a long/short portfolio if the short side is yielding a positive monthly return. I find the US stock market outperforms the Canadian stock market on average over the February 2010 to July 2017 period.

Fourthly I look at the relationship between liquidity, volatility and price discovery with short selling. I use the UK financial stocks short sale ban of the 2007-2009 financial crisis in order to explore the effects of short selling on liquidity, volatility and price discovery. I employ a control portfolio as well to see the effects of the short sale ban. I employ a GARCH model to explore the effects of short selling on volatility. I find that volatility is not affected during the short sale ban, this in turn questions the significance of short sale bans like many other studies before mine have done. I employ a Bid-Ask Spread Model to explore the effects of short selling on liquidity. I find that the Bid-Ask Spread Model shows good fit regression wise and that liquidity deteriorates during the short sale ban period. Lastly, I employ a Wald-Wolfowitz Runs Test to see fat tails in its distribution, this shows the effects of price discovery on a short sale ban. I find that price discovery deteriorates during the short sale ban period.

CONTENTS

RESEARCH AND EDUCATIONAL BACKGROUND

I come from a mathematical background with an interest in the stock market and investing. I hold a BSc Mathematics from the University of Greenwich and a MSc Finance and Econometrics from Queen Mary, University of London. I have also completed the first year of the PhD Economics programme at the University of Surrey where extensive study and research was taken in Advanced Macroeconomics, Advanced Microeconomics and Advanced Econometrics. Research wise further on I have conducted research on the volume-volatility relationship in the United States as a master's dissertation for the Queen Mary, University of London Finance and Econometrics programme.

I believe I have good knowledge in Investment Theory, Corporate Finance, Financial Derivatives, International Finance and Financial Econometrics which are important areas of study in production of a PhD of this nature. Thus, I believe I have the required background and knowledge to undertake a PhD in this field.

LIST OF ABBREVIATIONS

The following is a list of abbreviations used throughout this thesis. These abbreviations apply across the thesis unless stated otherwise. The meaning of these abbreviations is also listed throughout the thesis but are also written here as a primary reference in a scenario where parts of the thesis are removed for reference.

ADR – American Depository Receipt

AGARCH – Asymmetric Generalised Autoregressive Conditional Heteroscedasticity

AIC – Akaike Information Criterion

AIM – Alternative Investment Market

APT – Arbitrage Pricing Theory

AMEX – American Stock Exchange

ARCH – Autoregressive Conditional Heteroskedasticity

AR – Autoregressive

ARMA – Autoregressive Moving Average

BIC – Bayesian Information Criterion

BLUE – Best Linear Unbiased Estimator

CAPM – Capital Asset Pricing Model

CBOE VIX – Chicago Board Options Exchange Volatility Index

CFD – Contract for Difference

CMA – Conservative Minus Aggressive

CME – Cheap Minus Expensive

CRSP – Centre for Research in Security Prices

DJIA- Dow Jones Industrial Average

DMM – Designated Market Makers

EGARCH – Exponential Generalised Autoregressive Conditional Heteroscedasticity

EM-RF– Expectation of Market Return minus Risk Free Rate

EMH – Efficient Market Hypothesis

EPS – Earnings Per Share

ETF – Exchange Traded Fund

FSA – Financial Services Authority

FCA – Financial Conduct Authority

FOMO – Fear of Missing Out

FTSE – Financial Times Stock Exchange

GBP – Pound Sterling

GARCH – Generalised Autoregressive Conditional Heteroscedasticity

GDR – Global Depositary Receipt

GMM – Generalised Method of Moments

GLS – Generalised Least Squares

HML – High Minus Low

IPO – Initial Public Offering

ISA – Individual Savings Account

IT – Information Technology

IV – Instrumental Variables Regression

LHS – Left Hand Side

MA – Moving Average

MOM – Momentum

MLE – Maximum Likelihood Estimation

NASDAQ – National Association of Securities Dealers Automated Quotations

NYSE – New York Stock Exchange

OECD – Organisation for Economic Co-Operation and Development

OLS – Ordinary Least Squares Regression

P/B – Price to Book

P/E – Price to Earnings

PEG – Price to Earnings to Growth

PhD – Doctor of Philosophy

QGARCH – Quadratic Generalised Autoregressive Conditional Heteroscedasticity

RAM – Random Access Memory

RHS – Right Hand Side

RMW – Robust Minus Weak

RSS – Residual Sum of Squares

SIR – Short Interest Ratio

SEC – Securities and Exchange Commission

SMB – Small Minus Big

SML – Security Market Line

SSB – Short Sale Ban

SSE – Sum of Squared Errors

STEM – Science Technology Engineering Mathematics

S&P – Standard and Poor's

TARP – Troubled Asset Relief Programme

UK – United Kingdom

US – United States

USD- United States Dollar

VC – Vice Chancellor

WW2 – World War Two

LIST OF TABLES

The following is a list of tables used throughout the thesis for reference. Tables are listed in numerically ascending order. Sources for tables are written respectively under each table in the thesis. Tables listed with "A" in front of the table number are in the appendix, all other tables are in the main body of the thesis.

LIST OF FIGURES

The following is a list of figures used throughout this thesis for reference. Figures are listed in numerically ascending order. Sources for figures are written respectively under each figure in this thesis. Figures listed with "A" in front of the figure number are in the appendix, all other figures without an A are in the main body of the thesis.

CHAPTER 1: INTRODUCTION

"The Predictive Nature of Short Interest on Market Returns and the Effect of Short Selling on Volatility, Liquidity and Price Discovery with Some International Evidence", is a study that has been put forward to add further understanding to the asset pricing model literature, short interest literature and short selling literature. All of these topics are central to the field of finance which has been one of the most modern and rapidly developing areas of study in the 20th century. Finance has branched off economics, which is the field studying the discrepancies between the wealth of nations, to create a new field which has focused on how capital is raised for firms and the creation of secondary markets which arise from this. Without a means of raising capital, firms and the global economy would not be able to grow, showing the essential nature of finance in the modern world. Even though great emphasis is put on secondary markets such as the stock market, bond market and derivatives market (consisting of instruments that derive value based on underlying's such as bonds and stocks and time expiration), the initial offering of bonds and stocks are crucial for the functioning of the global economy.

This need to raise capital lead to the formation of institutions which facilitated the gathering of buyers and sellers, the most synonymous being Wall Street in New York City. Originally starting as a place where people gathered to deal, these institutions developed into global exchanges where the whole world could interact with each other in terms of securities markets. As time went on these meeting houses became more and more open to the general public, as opposed to places in the past where hidden deals were made behind closed doors. The increase in liquidity lead to more stability in prices and improved market efficiency, a seller would always be available for a buyer. The introduction of Designated Market Makers (DMM) [1] further contributed to market stability and efficiency. Regulation also increased and counterparty risk was dramatically reduced, the SEC (securities and exchange commission) in the US and FCA (financial conduct authority) in the UK are two of many major bodies enforcing regulation on financial markets. These centres arose around the world in New York City, London, Tokyo, Shanghai, Mumbai and Frankfurt underpinning the global market system that we have today, with trillions of US dollars of capital held in it. These centres along with the major world central banks that of the US Federal Reserve, Bank of England, Bank of Japan, the European Central Bank and the People's Bank of China form the core of the global financial system. The central banks acting as the lender of last resort[2] in times of crisis's and underpinning the robustness of the global financial system.

Even with such a robust and global financial system, cracks have appeared from time to time in particular during the 2007-2009 financial crisis, where the entire global financial system was on the verge of breaking down. Only a government bailout from central banks around the world saved the global financial system from a worldwide depression. The collapse of the banking system would have meant that ATMs would have

[1] Designated Market Makers (DMM) provide liquidity to securities by regularly having buy or sell orders in the order book. This means that securities remain as liquid as possible, where there will always be a buyer for every seller and a seller for every buyer. The NYSE provides dedicated market makers for its securities. This is particularly of use when a security has low volumes for trading.

[2] The lender of last resort is usually a central bank that serves to lend to financial institutions that are in distress, in the aim of protecting the wider economy. These financial institutions in distress may not be able to raise capital in the open market, as thus may turn to the central bank as a lender of last resort to protect the financial institution from collapsing.

ceased to function within a few days and bank accounts would be empty overnight, there was ample need to inject liquidity into the system. Much credit has been given to the US TARP programme in achieving this stimulus from the US Federal Reserve. This stresses the need for good academic finance research to eliminate such catastrophes in the future, as their effects are felt far and wide, both inside and outside the financial system. Knowing this, this thesis aims to add to the academic finance literature.

Going back to our three topics in mind, the asset pricing model literature aims to find models that can represent the fair rate of returns in markets. This includes models which use factors such as volatility, momentum and the outperformance of certain types of stocks over others to see where returns are generated. The short interest literature links the behaviour of short interest with market returns. This includes looking at the differences between heavily shorted stocks and lightly shorted stocks in returns. The short selling literature looks at how short sellers affect the market in terms of liquidity, price discovery and volatility. This includes investigating short sale bans and how certain types of stock behave with restrictions. Overall these topics are very much interlinked and academic finance research papers use elements from all these topics to form hypothesises and conclusions.

The purpose of this project is to investigate whether the findings of Boehmer et al. (2010) in regards to short interest and stock returns are robust to model specification, investment horizon and across the markets of different countries. Model specification being whether the findings are similar when different asset pricing models are used, investment horizon being if the findings are similar across different time periods and markets of different countries being if findings are similar in different countries. My project will aim to address these three criteria in my research questions. The project further aims to explore to what extent short sales affect market metrics like liquidity, price discovery, volatility and cross section of stock returns. Given these findings this project gives a view on market manipulation and the informed or uninformed nature of short sellers in the market. This project also touches on the efficient market hypothesis and whether short sellers have a negative or positive effect on market efficiency.

1.1 Why this Study Matters?

One of the most focal points of a study is not just what a study is investigating but also why this investigation matters in the overall field of research, in my case in the field of finance. Research is most useful when it fills in gaps of knowledge or leads to the development of other research that progresses the field that the research is being undertaken in. I strongly hope and anticipate that this study will achieve both of these objectives.

The first reason this study matters is because it will be one of the first studies to test an advanced asset pricing model (Fama and French Five Factor Model) on short interest data[3]. The Fama and French Five Factor Model is the joint work of Eugene Fama and Kenneth French after years of study and collaboration in asset pricing models. Both authors made significant contributions to the test of the CAPM (Capital Asset

[3] Short interest is the percentage of shares that has been sold short and still not covered in regards to the total number of shares able to be sold short in the market (often referred to as the float). Short interest is an indicator of future buying demand. Once a stock starts to move up in price, short positions may cover and increase buying pressure. This may cause a stock to move up even further in price, causing even more short positions to close. This is referred to as a short squeeze.

Pricing Model), developed the Fama and French Three Factor Model and subsequently went on to publish their Fama and French Five Factor Model.

From this study we will be able to see firstly the ability of the Fama and French Five Factor Model to give fair asset returns and also whether the findings of Boehmer et al. (2010) are robust to model specification. A study that is conducted often needs robustness checks to see if their findings are consistent. Boehmer et al. (2010) focused their entire study on the Fama and French Three Factor Model augmented with Momentum, often referred to in the literature as the Carhart Four Factor Model. In their study Boehmer et al. (2010) found that there was indeed good news in short interest, that is an excess return could be achieved by going long a portfolio with the least shorted stocks and short a portfolio with the most shorted stocks and rebalancing each month to consider new short interest data. This means buying the stocks with the lowest short interest and selling the stocks with the highest short interest and replicating this procedure every month to account for new short interest data. This short interest data is now is currently available twice a month. It must also be added that this was achieved on a risk adjusted basis using the Fama and French Three Factor Model augmented with Momentum[4] as their adjuster for risk.

Secondly this study has implications for asset managers and central banks as a whole. My study looks at effectively a way of generating a positive alpha and these forms of mechanical strategies are liked by fund managers to manage their clients' money. Mechanical strategies are those that require little active input and can in turn be automated by a computer algorithm on when to buy and sell a security, this often removes the daily stress of watching markets closely. By going long, a portfolio of lightly shorted stocks and going short a portfolio of heavily shorted stocks and rebalancing each month, Boehmer et al. (2010) showed a way of generating a positive alpha[5]. I am looking at whether this strategy employed by fund managers is still effective today as well or whether it has been arbitraged out. An efficient market would arbitrage out such a strategy, so this is also a test of market efficiency and my findings can be used for the Efficient Market Hypothesis (EMH) debate. This is thus a good benchmark for fund managers to understand whether the strategy of Boehmer et al. (2010) is still applicable in the market today.

My study documents international evidence as well, again adding to the robustness of Boehmer et al. (2010). Many studies such as Desai et al. (2002), Asquith et al. (2005) and Boehmer et al. (2010) are focused on US Markets as a whole and it is very important to look at the global macro environment rather than a region-specific environment when evaluating asset return strategies. Strategies employable in one country may not be applicable in another, due to differences in volatility, volume and restrictions on the options market and short selling.

This also allows us to see whether short sellers are informed or uninformed and thus whether they engage in market manipulation. Being informed means trading around the news available, while being uninformed means trading with little to no information as guidance. This again puts into question the role of short sellers in the market and the policies central banks should consider regarding banning or implementing their

4 Momentum is factor in an asset pricing model based on holding stocks which are past winners over the previous month over past losers, as stocks with positive momentum are seen to keep rising in low volatility market conditions.

5 Alpha is the over or under performance of a portfolio in relation to a model benchmark, a positive alpha is a portfolio which is performing better than its model benchmark anticipates.

participation in markets. If the majority of short sellers are seen to be manipulating the market for their own monetary gain at the expense of others, regulatory authorities would be against their inclusion in the market. Short sellers can manipulate the market with several strategies, they can spread false rumours to depress the price of a stock, or engage in predatory activity where multiple short sellers target the selling of a certain stock in order to depress the price.

Central banks are often interested in implicating policy that is stabilising to the markets, usually on the monetary side of economic operations. Stabilising falling stock markets is often a priority of market regulators and can help to bring confidence back into a market. Stock exchanges often stabilise markets with circuit breakers, where trading is halted if the market or a particular security falls a certain percentage amount in value in a given trading day. Circuit breakers are these automated computer systems that are used to implement these trading halts. The NYSE for example has three levels of circuit breakers, Level 1, Level 2 and Level 3. Level 1 and Level 2 are activated at 7% and 13% respectively for drops in the level of the S&P 500 and result in a 15-minute suspension to trading unless the drop occurs after 3:25 pm, where trading is allowed to continue till the end of the session (4 pm). Level 3 is activated on a 20% drop in the S&P 500 and immediately suspends trading for the day.

In periods of extreme volatility, stock exchanges often open securities for trading one at a time in order to control large amounts of selling volume entering the market. It also allows market participants to see price discovery earlier than normal, so that panic selling may be lessened to an extent. Selling often brings on further selling pressure from stop losses being triggered and investors losing confidence in the market. Confidence is a key driver of markets, where a loss in confidence can cause a run[6] in securities or funds, similar to runs that have happened in banks in the past, in particular before the formation of central banks. Funds of poorly performing managers are in particular susceptible to runs and can often be suspended to stop this in the short term.

Central banks on the other hand can't control specific market mechanisms but can implement policy in order to stabilise markets. Central banks are often interested in mechanisms to reduce market volatility and it is often reflected as to whether short sellers should be banned in periods of severe market volatility. In this study I will also be exploring how short sales affect liquidity, price discovery, volatility and the cross section of stock returns. My study will shed light on this matter and give central banks an idea of how volatility is affected with the exclusion of short sellers.

1.2 Research Aims and Objectives

The purpose of this project is to investigate whether the findings of Boehmer et al. (2010) are robust to model specification, investment horizon and across the markets of different countries. In particular this project seeks to answer the following core four questions:

[6] A run is where depositors or investors simultaneously try to withdraw money out at the same time, this can cause the collapse in an institution or the fire sale of assets. This can be destabilising for the institutions involved. Runs can be prevented by restoring confidence in the market or bank and/or limiting withdrawals.

1. In adjusting for risk, Boehmer et al. (2010) uses the Fama and French (1993) three-factor model augmented by the momentum factor. However, is this long/short strategy still valid if a different and more recent model, such as Fama and French (2015) five-factor model, is used to adjust for risk premium? This question will be referred to as "research question 1" throughout the study.
2. After the publication of Boehmer et al. (2010), does the opportunity for excess returns remain, or have investors adopted this strategy and arbitraged away the excess returns? This question will be referred to as "research question 2" throughout the study.
3. Is the long/short strategy of Boehmer et al. (2010) still valid in an international OECD market that of Canada? This question will be referred to as "research question 3" throughout the study.
4. Whether and to what extent short sales affect liquidity, price discovery, volatility and cross-section of stock returns? This question will be referred to as "research question 4" throughout the study.

I will throughout the thesis refer to each question when appropriate and will structure my thesis around these questions. My objective of this thesis is to provide a conclusive answer to each of these four questions. The first three questions are very much interrelated as they work on the same back testing of portfolios and asset pricing model regression methodology, it would be good to see a comparison between the results of these three questions. The last question is unique in its methodologies. It will give a broader outlook on how short selling affects the market, rather than just on a risk adjusted returns basis (which research question 1 to research question 3 hope to achieve). Research question 4 also involves several different methodologies such as of a GARCH model, a bid-ask spread model and a runs test that aim to target volatility, liquidity and price discovery respectively. These three metrics encompass most relevant changes in the behaviour of a stock that is of interest to investors, speculators and central banks as a whole.

1.3 Software and Computational Requirements

The software and computational requirements for this project are in line with most finance and econometrics research that uses quantitative data. Research questions 1,2 and 3 will require Microsoft Excel to compile data and form portfolios. STATA is also required for research questions 1,2,3 to run regressions. Research question 4 will require STATA to run models for volatility and liquidity. In particular GARCH family of models can only be run in econometric software such as STATA and EViews, though third-party extensions for Microsoft Excel do exist for advanced econometric models. Microsoft Excel will also be required for research question 4 in order to compile data.

Data collection wise, Thomson Reuters DataStream[7] will be used to collect stock return data, trading volume

[7] Thomson Reuters DataStream is an online database run by Thomson Reuters that provides financial data to banks, universities and other financial institutions for the means of research. The database is updated regularly. Database can be accessed via subscription at: https://eikon.thomsonreuters.com/index.html

and high/low price data. Fama and French Factor data for the factors of beta, SMB, HML, CMA, RMW, MOM and the risk-free rate data will be collected from Kenneth's French's database, which is available online[8]. The factors are compiled from portfolios of stocks which hold those respective characteristics.

Computationally files can be very large and will require a good deal of memory on any computer used, especially in the case of excel files with over 100+ spreadsheets for the portfolios formed from short interest. Therefore, a computer with an adequate amount of processing power and random-access memory (RAM) will be required. Maximum Likelihood Estimation can also be computationally taxing for a computer, so again an adequate amount of processing power will be required for that. Maximum Likelihood Estimation requires iterations, which can run infinitely if there is a flat log likelihood, therefore requiring computer processing power.

Microsoft Word will be used for this thesis, again file size needs to be considered as pictures in a document with over 100 pages can be very large and can exceed 100 MB. This means sending files across email can be of issue, though this can be addressed by shrinking files using PDF and/or sending documents via the postal system. LaTex is an alternative software package that can be used, in particular if the document is equation heavy. LaTex is usually the preference in STEM fields (in particular Mathematics), however Microsoft Word has added flexibility and compatibility with external software such as STATA and Adobe PDF. Overall the software and computational requirements of this thesis are adequately met by the resources available to me both at the university and at home.

1.4 Summary Empirical Results

I report full results for each of the research questions separately and make comparisons when necessary in the results chapter of this thesis. For a full analysis of results, the readers of this PhD thesis should look at the results of this thesis in Chapter 6. Below I list summary empirical results for ease of reference.

Regarding research question 1, I look at whether the strategy of Boehmer et al. (2010) is still valid for a new and more efficient

model such as the Fama and French Five Factor Model. The original research was conducted using the Fama and French Three Factor Model with Momentum. I find that the most heavily shorted stocks underperform lightly shorted stocks. The Fama and French Five Factor model holds a higher positive alpha on both lightly shorted portfolios compared to their heavily shorted counterparts. The alpha on the lightly shorted portfolios being 0.025 and 0.019 for the SIR 5% and SIR 10% portfolios respectively compared to the alpha of 0 for the SIR 95% and SIR 90% portfolios respectively. The alpha represents the intercepts of the regression, and accounts for the over and under performance of the asset pricing model.[9] Both lightly shorted portfolios hold a much higher raw return of 3.1% and 2.6% for the SIR 5% and SIR 10% portfolios respectively

[8] Kenneth French's database can be accessed at: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[9] For example, if the expected return from an asset pricing model was 2% a month and the portfolio returned 3% a month, we would say that the alpha of the portfolio was 1% a month.

compared to 1.9% and 1.5% for the SIR 95% and SIR 90% portfolios respectively. There is good news in short interest with the Fama and French Five Factor Model, however I suggest going long the SIR 5% portfolio as no heavily shorted portfolio holds a negative raw or excess return.

For research question 2, I look at whether after the publication of Boehmer et al. (2010) arbitrage has taken place or not and whether the strategy of going long the least shorted stocks and going short the most shorted stocks is valid. I find that the most heavily shorted stocks underperform the most lightly shorted stocks by 0.9% a month on a raw return basis, this is consistent across my findings, with the underperformance of heavily shorted stocks. The original strategy by Boehmer et al. (2010) was to go long the top percentile portfolio and go short the bottom percentile portfolio. I find that my SIR 95% portfolio yields a positive average return of 1.7% a month, therefore it is not advisable to have a short component in your strategy. My findings suggest that it is best to go long the SIR 5% portfolio, as this provides the best raw return and risk adjusted alpha out of all the portfolios.

Regarding research question 3, I look at whether the findings of Boehmer et al. (2010) are applicable to another country marked as an OECD. For this question, I use a country like Canada as a proxy for the OECD label, given the vast availability of short interest data from the Toronto Stock Exchange. Short Interest Data can be difficult to obtain[10] for smaller markets across large datasets. My results indicate again that heavily shorted stocks underperform lightly shorted stocks. I find that the most heavily shorted stocks underperform the most lightly shorted stocks by 1.5% a month on a raw return basis. This is consistent with my findings in research question 2 and consist with the previous literature as a whole. However again I find that my SIR 95% portfolio yields a positive raw return of 1% a month on average, this indicates that it is not viable to have a strategy that includes a short component. I again suggest the best strategy is to go long the least shorted stocks and therefore go long the SIR 5% portfolio.

For research question 4, I look at whether and to what extent short selling affects volatility, liquidity and price discovery in markets. I use the short sale ban of 2008 in UK financial equities as a period of investigation, as the only means of investing the effects of short selling are times during which short selling has been banned as in most market conditions short selling is in place in non-emerging markets. My results indicate that volatility remains similar across the short sale ban and no short sale ban periods, however liquidity and price discovery deteriorate during the short sale ban period. My findings for volatility, liquidity and price discovery are consistent with the existing literature.

1.5 Structure of this Thesis

The rest of this thesis is split into the following chapters, Chapter 1 is the Introduction, Chapter 2 is the Literature Review, Chapter 3 is the Theoretical Framework, Chapter 4 is the Data Description, Chapter 5 is the Estimation Procedures, Chapter 6 is the Empirical Results and Chapter 7 is the Conclusion. This makes

---

10 From my data provider of Thomson Reuters DataStream, short interest data was only available for North American Markets (US and Canada). I would have preferred to explore markets outside of North America, but there is a lack of data on short interest on all European Countries. Thomson Reuters DataStream is aware of this and is trying to expand its datasets in short interest across the major equity markets of the world.

up the main body of my thesis. References follow Chapter 7. The chapters are divided into further sections to break up material and allow the reader to understand how what is being explained relates to the main section. The appendix of this thesis contains extra information relevant to the project that is not essential in the main body of text. The appendix includes key tables and diagrams that are of interest and may be useful to see in relation to my main result findings.

Chapter 1 is the Introduction and will outline why this study is important in terms of asset managers, central banks and regulatory authorities and why I choose to do this study over something else. Chapter 1 also outlines my research aims and objectives in the study and what I hope to achieve once the study is over. Chapter 1 will also show the empirical results of what I found and will lay out the structure of this thesis.

Chapter 2 is concerned with the Literature Review. This is an in depth look at the current literature in both short interest and asset pricing models. I am in particular concerned in studies done on short interest and/or short selling that involve market manipulation, stock returns, volatility, liquidity, price discovery and political aspects.

Chapter 3 is the Theoretical Framework, and I will look at the composition of the models that I am using such as the CAPM, Fama and French Three Factor Model, Fama and French Five Factor Model, GARCH Model and Bid-Ask Spread Model. I also look at the mathematics behind them and their derivation from source. I will also look at the theories in finance I am employing such as the Information Hypothesis, Overvaluation Hypothesis and Informed Traders Hypothesis and the background behind them. This section very much interlinks with the literature review to give a broad background to the research questions.

Chapter 4 is concerned with the Data Description. This chapter outlines the data I am employing in the form of monthly short interest ratio, individual stock returns and trading volume data from Thomson Reuters DataStream and Fama and French Factor Data from Kenneth French's Data Library. This gives a clear outlook of where my data came from and adds to the replicability of my study. I also give data statistics in the form of short interest ratio graphs, returns graphs and trading volume graphs to give a better view of the datasets used in each research question.

Chapter 5 is my Estimation Procedures. This chapter outlines the methodology I employ such as a Calendar Time Portfolio technique with Ordinary Least Squares Regression, Maximum Likelihood Estimation with GARCH, Ordinary Least Squares Regression with a Bid-Ask Spread Model and a Runs Test in order to achieve my results. I also explain why I have chosen these methodologies over others. I also outline the consistency of my methodology econometrically, which is crucial to get good results.

Chapter 6 is my Empirical Results. This chapter deals with what I have found in terms of performance of certain short interest portfolios over others and the impact of a short sale ban on volatility, liquidity and price discovery. This section tries to tie up my mathematical findings to my theory in order to draw conclusions on what the data might be telling us. I present full tables and analysis in this section.

Chapter 7 is my Conclusion based on my results. It also outlines what I can do to take this study further and what the answers to the original research questions set put in my introduction are. The rest of the thesis concludes with references that show my sources and other information that adds to the main body of this thesis. This outlines the structure of my thesis.

1.6 Conclusion

This concludes the introduction chapter of this thesis. We have seen the origins of finance and how it is important to the functioning of modern-day society. The need to raise capital in public markets naturally lead to the formation of stock exchanges. The need of a lender of last resort naturally lead to the formation of central banks. We have also seen the importance of the study I am conducting and the implications my research has on a wider context. A large number of people will benefit from a study like this including asset managers and central bankers, this study will also add further depth into the broad literature on asset pricing models, short selling and short interest.

I have also outlined my research aims and objectives. I seek to answer four questions in particular and will build my thesis around these questions. All questions seek to analyse the predictive nature of short interest and returns and also the behavioural nature of short sellers in the market.

I also list the software and computational requirements of such a study, in particular noting the use of Microsoft Excel and STATA for data manipulation and model use respectively. I use Microsoft Word as my core document software, noting the benefit of LaTex for STEM fields.

We have seen the empirical results of my study and have linked them to my initial questions and whether my hypotheses' hold true or not. Further analysis of my empirical results can be found in the results chapter of this thesis.

We have seen the complete structure of the thesis, which is clearly defined into chapters and sections depending on the question being answered. I follow a very standard procedure in the layout in regard to thesis writing. I also include footnotes where applicable to add further depth and clarity to information. These footnotes are indicated with numbers respectively on the top of right of text in the main body. Each chapter includes a conclusion such as this which summarises findings.

CHAPTER 2: LITERATURE REVIEW

It is of great essence to conduct a good theoretical framework and literature review in the topic of your study. This chapter aims to outline the literature review. The literature review deals with the past and current literature and is aiming to bring the reader up to speed with the current state of play. The theoretical framework is not concerned with individual papers, but more concerned with the finance theory in hand applicable to the research. The theoretical framework will follow the literature review.

The literature review of this thesis is not limited to one topic of research in finance, as has been made evident by the research questions posed. The first area in my literature review I must develop is that of asset pricing models. The first three research questions of my research rely heavily on asset pricing models and this literature needs to be assessed to understand where we stand currently.

The second topic that I must develop is the central topic in this research, is that of short interest and short selling. I need to see what the current literature on this topic is in relation to returns in the stock market. I will also look at short selling and the impacts that it has on the market and market returns. Short selling is related to short interest and the short interest ratio, but is different in the fact that in short selling the positions are closed, while with short interest the positions are open. The short interest ratio is a ratio which compares the number of open short positions with the total number of stock available for shorting. I will also be looking at how short selling affects liquidity, volatility and price discovery in stocks and whether we have informed traders and thus if there is market manipulation[11] in short selling. I will in addition explore the political aspects of short selling and the case of whether short sellers are moral individuals. These two topics I have mentioned will form the core of my literature review. This will be the starting point on which I will be able to answer my research questions.

The two topics of asset pricing models and short selling have a very long history in finance and fields outside of finance. Asset pricing models have been in development around after the 1950s with the work of Nobel Prize winner Harry Markowitz and his Mean Variance Theory. Short selling is seen to be around since 1609 and is said to be attributed to Dutch businessman Isaac Le Maire (Kalse and de Graaf, 2008). In regards to financial markets, short selling has been around since the foundations of the Amsterdam Stock Exchange and now the more popular New York Stock Exchange. Short sellers were in particular blamed for the Wall Street Crash of 1929 (Reuters, 2008), that still to this day is the single largest drop in US stock market history.[12] The perception of short sellers is often seen negatively by investors and often regarded as a cause of extra downward pressure in a market. The short interest ratio as a gauge is as deep as short selling itself as it shows the number of open short positions.

---

[11] Market manipulation in short selling refers to any activity that benefits short sellers in making a profit that is against the laws of regulatory authorities like the SEC and FCA. Short sellers may spread false rumours and engage in predatory group targeting to bring down the price of a stock.

[12] The Wall Street Crash of 1929 took out nearly 90% of the market capitalisation of the Dow Jones Industrial Average and signalled the start of the Great Depression. The DJIA took almost 25 years to recover from the peak of the crash, this may seem extreme compared to recent recessions but the US had entered a depression which saw extremely high levels of unemployment (over 25% in some cases) and had experienced a great deal of deflation (Reuters, 2008).

2.1 Asset Pricing Models

The central topic of interest in portfolio management has always been that of risk and return and how to quantify the two meaningfully. The job of asset pricing models has always been to judge the fair return of a security given the rate of return of other securities considering individual and macro risks of the market as a whole. The literature in asset pricing models is vast and huge developments have been made over the last 70 years such as the Capital Asset Pricing Model, Arbitrage Pricing Theory and Multifactor Arbitrage Pricing Theory Models.

This topic is of significance to us as I am using two relatively new asset pricing models in my research, that of the Fama and French Five Factor Model and of the Fama and French Three Factor Model with Momentum. The Fama and French Five Factor Model is such a recent development that it was in fact published after one of my core reference papers for this PhD, Boehmer et al. (2010), so it puts us on the cutting edge of research and development in finance.

2.1.1 Capital Asset Pricing Model

The majority of the 1950s to 1970s built on the Capital Asset Pricing Model in finding a fair return for a security given its risk. The later years were dedicated to the testing the model and making refinements on it. This was one of the major achievements in asset pricing and papers still to this day are written on the consistencies of the model and its applications.

Markowitz (1952) is one of the early papers where the initial foundations of asset pricing models begins to take place. Markowitz (1952) introduces us to the concept of maximising expectation of returns and minimising variance of returns. These are the two priorities any portfolio should be aiming for and would be the benchmark of a rational investor. Markowitz (1952) introduces the efficient frontier, a set of portfolios that a rational investor would hold. He also notes that the variance of a portfolio increases as the covariance between securities increases, therefore it is advisable to hold securities which are not highly correlated with each other. Markowitz (1952) lays the initial foundations for the Capital Asset Pricing Model in this paper.

Modigliani and Miller (1958) is the other paper that lead to the development of the Capital Asset Pricing Model. Modigliani and Miller (1958) explore the connection between a firm's capital structure and the cost of capital, finding that a levered and unlevered firm has the same value. The inability to determine the correct discount rate contributed to Jack Treynor in his analysis for it and thus the development of the Capital Asset Pricing Model.

Subsequently, Sharpe (1963) also creates a model known as the Single Factor Model in pre-Capital Asset Pricing Model development. This model states the return of a security can also be decomposed to its expected plus unexpected components, the unexpected components can be further decomposed to macro and firm specific uncertainties. A sensitivity coefficient is also applied to the macro specific uncertainties.

Treynor (1961,1962), Sharpe (1964), Lintner (1965) and Mossin (1966) working independently lay down what is known as the Capital Asset Pricing Model. A model used to determine the expected return of a security, given the risk-free rate, beta[13] of the security and the excess return of the market[14]. The model states that the expected return is a function of the risk-free rate plus the volatility of the security in relation to the market (beta) multiplied by the excess return of the market.

Building on, Black (1972) shows that the major results of the Capital Asset Pricing Model (CAPM hereafter) does not require a risk-free rate. Investors instead use a zero-beta portfolio, a portfolio of risky assets that has zero covariance with the market portfolio. Thus, the Zero Beta CAPM is developed using beta as a measure of systematic risk.[15] Merton (1973) develops the ICAPM or Intertemporal CAPM. In the ICAPM the single time period assumption of the CAPM is relaxed and trading thus takes place continuously in time. The ICAPM moves from discrete time finance to continuous time finance.

Further on, Roll (1977) publishes his famous Roll's critique, an analysis of the validity of the empirical tests on the CAPM. Two statements are made regarding the market portfolio. The first being any mean-variance efficient portfolio satisfies the CAPM exactly and the second being that the market portfolio is unobservable.[16] If the market portfolio is unobservable it is impossible to test the CAPM. Stambaugh (1982) develops on what Roll (1977) has proposed and expands the market portfolio to include corporate and governmental bonds, consumer durables and real estate. He finds the conclusions from the CAPM tests are not affected as investors increase the size of their market proxy (item used to represent the market portfolio). It is worth noting that all these assets are tradable.

In the same year, Basu (1977) poses another blow to the CAPM, who finds evidence that when stocks are sorted on a basis of the price to earnings ratio[17] (P/E hereafter) future returns on low P/E stocks are higher than the fair return estimated by the CAPM.

Banz (1981) says that when stocks are sorted by market capitalisation (often taken as the last traded market price of a security multiplied by the number of outstanding shares) the average returns experienced by small cap stocks are higher than the fair return for the level of risk being taken given by the CAPM model.

Bhandari (1988) also finds that companies that have high levels of leverage, defined as the book value of debt over the market value of equity, are getting returns that are higher than expected given the beta of the

---

13 The beta of a stock is the volatility of the stock with respect to the market portfolio, a stock with a historic beta over 1 has been more volatile than the market. A stock with a beta less than 1 and above 0 has been less volatile than the market. Beta can also be negative and so hold inverse volatility with the market.

14 The excess return on market is calculated by taking the return of the market portfolio and subtracting it from the risk-free rate. This gives the market risk premium for holding the market portfolio over a risk-free rate.

15 Systematic risk is the risk that can't be diversified away from holding more securities, thus a higher return is required to compensate for it by the rational investor.

16 The market portfolio is seen to be unobservable, as the true market portfolio has to include every single possible asset of value of which some have unobservable returns such as assets not listed as marketable securities.

17 The price to earnings ratio is often used as a gauge for how expensive or cheap a security is. When a security is trading for a low price to earnings ratio market expectations for earnings growth are lower than for securities trading on higher price to earnings ratios. Investors are paying today for the future earnings ability of the security as the market is always forward looking. The price to earnings to growth ratio (PEG ratio) is often a better indicator of the overpricing or underpricing of a security as the P/E ratio should equal the growth in earnings per share in a fairly price security. Therefore a PEG ratio of 1 indicates fair pricing for a security.

companies. By this time, the CAPM is starting to show structural problems and more and more effort is going into finding a model which better fits the return data.

A few years later, Jagannathan and Wang (1996) propose the Conditional CAPM. Empirical studies of the static standard CAPM model assume that betas remain consistent over time and that the return of the value weighted portfolio is a proxy for the average investor's portfolio. Jagannathan and Wang (1996) assume that betas and market risk premium vary over time, their new specification holds well in explaining the cross section of average returns in the long run, where most firms do as well as the market return, but there are a few extreme outliers driving the market positively. However, in the short run, results are not as good.

Statistically, Harvey and Graham (2001) find that the CAPM is by far the most popular method for estimating the cost of equity capital in firms, with 73.5% of respondents using it. This may be due to the relative simplicity of the model and that it compares two important factors in finance, that of risk and return.

More recently, Fama and French (2004) review the evidence and theory of the CAPM. They argue that the model is still a very effective way of estimating the cost of capital, even with its drawbacks.[18] Fama and French (2004) say that some of the CAPM assumptions are so fundamental that if they were ignored many other models would become redundant as well. In particular the assumption of the market proxy, all asset pricing models require a good market proxy and saying investors' can't measure the market would make 99% of asset pricing models redundant. There must a comparison from the return of the portfolio to a fair return for that portfolio in some means or another.

This gives a brief overview of the developments of the CAPM and how still in modern times it is under scrutiny, though it is a very important development as it is the foundation of most asset pricing models.

2.1.2 Arbitrage Pricing Theory

Building on from the CAPM, we have Arbitrage Pricing Theory (APT). Unlike the CAPM model, which assumes that markets are perfectly efficient, APT sometimes misprice securities, before the market eventually corrects and securities move back to fair value. Arbitragers hope to take advantage of any deviations from fair market value. This not a risk-free operation in the classic sense of arbitrage, because investors are assuming that the model is correct and making directional trades, rather than locking in risk-free profits.

Ross (1976) proposes the APT, which stresses that the returns of securities can be described by a factor model, liquid and well-regulated financial markets will not allow arbitrage opportunities and that there exist enough securities to diversify idiosyncratic risk.

[18] Drawbacks of the CAPM are ingrained in the assumptions of the model, such as perfect market competition, all investors being able to lend and borrow at the risk-free rate and the ability to measure the market return. A lack of efficiency in a market, can amplify the drawbacks of the CAPM.

Adding on, Ross and Roll (1980) conduct an empirical investigation on the APT theory. They use data for individual equities during the 1962 to 1972 period on American exchanges and find that at least three and probably four factors are responsible for generating returns.

A few years later, Lehmann and Modest (1985) provide an examination of the various strategies of constructing portfolios that are correlated to a large degree with the factors are affecting security returns. Three main conclusions are established from their study. The first one being increasing the number of securities in the portfolio improves the performance of the portfolio when this correlation is present. Secondly, they find that MLE (Maximum Likelihood Estimation) outperform other estimation procedures. Thirdly their Minimum Idiosyncratic Risk Portfolio Formation Procedure[19] outperformed the Fama-MacBeth estimation procedure.

2.1.3 Fama and French Factor Models

The work on the CAPM and APT theories lead to the development of larger and more complex models incorporating more factors that drive returns. The reason is that the CAPM laid down the foundations of giving a fair rate of return given the risk a market participant is taking, but many debates continued on the true components of that return.

Initially, Chen, Roll and Ross (1986) examine five macro variables for an asset pricing model. The factors being changes in expected inflation, unanticipated inflation, monthly growth rate in industrial production, unanticipated change in risk premium and unanticipated change in term structure. Chen, Roll and Ross (1986) find that industrial production, risk premium on corporate bonds and unanticipated inflation are statistically significant factors to explain returns.

Subsequently, Fama and French (1993) build on the work of the CAPM to develop a model known as the Three Factor Model in order to better describe asset returns. The Three Factors in the model being market risk (beta multiplied by the excess return of the market), price to book ratio and company size. Fama and French (1993) state that two classes of stock have seen to do better than the average, small cap stocks and stocks with low price to book ratios.[20]

A few years later, Carhart (1997) subsequently extends the Three Factor Model to a Four Factor Model accounting for momentum as the fourth factor. The momentum is taken as being long the winners of the prior month and short the losers of the prior month, a short time frame that is very evident in behavioural finance. The Carhart Four Factor Model is often used as the benchmark for active fund managers.[21] The

---

19 Fama-MacBeth portfolios are the sample minimum idiosyncratic risk portfolios which have a loading of one on one factor and a loading of zero on the other factors. Minimum idiosyncratic risk portfolios are constructed to have a sample loading of zero on the same factors.

20 These stocks with low P/B ratios are often referred to as value stocks, as opposed to growth stocks. The market has generally priced in lower growth for these stocks, so the market will be much more forgiving with earnings and guidance misses when these companies report quarterly results.

21 Active fund managers are seen to be ones who are prepared to hold a portfolio that changes consistently in order to produce an excess return of the market. Passive fund managers are seen to either hold the market portfolio or hold a portfolio with very little trading.

Three Factor Model and the Carhart Four Factor Model are often viewed as specific multi factor APT models[22].

Building on, Griffin (2002) subsequently shows that the Fama French factors on the Three Factor Model are indeed country specific to the UK, US, Japan and Canada. A conclusion is proposed that local factors may provide a better guidance to model stock returns than global factors. The aforementioned countries are some of the most commonly studied with factor models due to the large liquid markets and depth of information freely available.

More recently, Foye, Mramor and Pahor (2013) put forward an alternative factor model with three components. They show that the market risk premium factor becomes redundant when the Fama and French Three Factor Model is applied to emerging markets. This could very well be due to a flatter beta in emerging markets compared to developed markets. The alternative model proposed by Foye, Mramor and Pahor (2013) replaces the market risk premium with a term that acts as a proxy for accounting manipulation. Emerging markets are changing at a greater rate compared to developed markets, so countries classified as emerging markets need to be amended over time.

Novy-Marx (2013) finds that profitability of a firm (taken as revenues minus costs) has roughly the same power as book-to-market in predicting the cross section of average returns. This is one of the papers that influences Fama and French to adjust their Three Factor Model to include the profitability premium.

Thus, Fama and French (2015) extend their Three Factor Model further to a Five Factor Model. The two factors that are further added are that of profitability and investment. The Five Factor Model is seen to perform better than the Three Factor Model as judged by the Gibbons, Ross and Shanken (1989) test. This tests whether the factors fully explain the expected returns of portfolios. The failure of the Five Factor Model often comes from portfolios that have small firms that invest despite low profitability. The alpha of the model would be zero, if the model could explain all excess returns, as the alpha represents the unexplained returns. There are also arguments that state more factors may not necessarily be better, as it increases the chance of component error in the study and brings in the possibility that terms may be highly correlated.

Drechsler and Drechsler (2016) extend the Carhart Four Factor Model to include a CME factor. CME represents Cheap Minus Expensive and explains the return premium of high short-fee stocks over low short-fee stocks[23]. They say their model is a great extension over the Carhart Four Factor Model.

Building on from this, Chiah et al. (2016) test the Fama and French Five Factor Model in pricing Australian equities. They find that the Five Factor Model holds better than other competing asset pricing model in explaining anomalies, thus holding it as a superior model. Chiah et al. (2016) also find that in contrast to Fama and French (2015), the book to market factor retains its explanatory power when used in combination

22 Multifactor models may use various factors to explain returns, in the future there may be models which better explain returns or the data may change to make our current models redundant.

23 Stocks can be shorted and there is usually a fee associated with shorting them, some stocks are more expensive to short relative to others. Stocks which are seen by the market as overvalued are more expensive to short than stocks seen by the market as undervalued. The fee of shorting stocks comes from the fact stocks must be borrowed to be shorted unless you are shorting without borrowing (referred to as a naked short sale) or dealing with derivatives to form a short position. A natural derivative position on the short side would be a put option with a long-dated expiry. However, put options have an expiry and a new put option must be bought once the old put option expires if a short position is still to be held.

with the investment and profitability factors. Section 3.1.3 explains the construction of the investment and profitability factors in more detail.

Also, Lin (2017) tests the Fama and French Five Factor Model in the Chinese equity market. They find the Five Factor Model consistently outperforms the Three Factor Model in the Chinese equity market. Their findings are in contrast to Fama and French (2015), as Lin (2017) finds that the value and profitability factors are important, while the investment factor is redundant for this Chinese equity market sample.

Ulku (2017) explores the Robust Minus Weak (RMW) factor of the Fama and French Five Factor Model. RMW being the profitability factor of a firm, comparing low and high profitability firms. Ulku (2017) finds that the profitability factor is an early in the week effect that seems to have the greatest factor premium on Monday or Tuesday. This may reflect investors becoming more rational [24] over a weekend trading break.

Concerns have been raised about the effectiveness of the Fama and French Five Factor Model. For instance, Blitz et al. (2018) poses five concerns with the Fama and French Five Factor Model. Blitz et al. (2018) maintains that the CAPM's relation between market beta and return is too flat or even negative questioning the significance of that factor in the Fama and French Five Factor Model. Secondly, the momentum effect of the Carhart Four Factor Model is ignored and seen to be somewhat widely accepted as a basis for returns, especially with the development of behavioural finance over the last twenty years. Thirdly robustness concerns have been seen for the two new factors of profitability and investment, with the addition not being explanatory for returns. Fourthly the economic rational is not clear for the two new factors, where risk-based explanations were given for the Fama and French Three Factor Model. It seems the Fama and French Five Factor Model is not going to lead to a consensus view among the academic finance community. It may be the best asset pricing model we currently have, but many are not satisfied, especially with using the CAPM as a base of the model with all the strict assumptions the CAPM brings.

2.2 Short Interest and Short Selling

The short interest and short selling literature do not have the breadth and depth that the asset pricing literature has. The literature on short interest and short selling has focused mainly on periods of short selling constraints such as the 2007 to 2009 global financial crisis. The previous time to this that short selling was banned was in September 1931, when the NYSE banned short sales in the announcement that England was abandoning the gold standard. This was the only short sale ban in the US pre the financial crisis of 2007 to 2009.

Short interest and short selling are both very interlinked topics in the sense that short interest is a measure of open short positions and short selling is the act of opening and closing a position. Short interest is often reported in the form of either open short interest, the number of short positions open on the security or

[24] There is considerable debate in finance over whether the profitability factor of the Fama and French Five Factor Model is a behavioural mispricing or a rationally priced risk. This paper helps contribute to the debate.

that of the short interest ratio (SIR), the number of short positions in the security relative to the number of shares that can be shorted in the market[25]. Both metrics are used as indicators for shorting.

The process of short selling is the opposite the process of going long or buying a security, an investor gains money when a security drops in value as opposed to rising. To go short a security, an investor would need to borrow a security, sell it, buy it back at a future date and return the security. It is also possible in certain cases to sell short naked, where an investor is short without borrowing the security. With the advent of the derivatives market, an investor is able to go short without evening borrowing a security, put option contracts mean an investor can take a short position with a counterparty without even owning or borrowing the security. A Put option gives the holder the right to sell at a pre-determined strike price when the contract expires or before expiry depending on the conditions of the put option (American and European Put Options vary on their conditions of sale).

One of the major disadvantages of short selling is the ability for losses to exceed the initial capital deposited, brokers will often force a margin call if this situation arises. In a margin call an investor or speculator is forced to raise further capital to keep a losing short position open. The benefit of put options over short selling is the exact opposite, where only 100% of the principal can be lost and no more.

2.2.1 Short Interest, Short Selling and Market Returns

A great deal of the short interest and short selling literature has been concerned with returns, whether there is a viable strategy by investing with short interest information. Short interest as a metric is readily available and gives market participants a view of future demand. It also allows market participants to see the risk they are taking when trading a security, as a security with a high future demand is riskier to short.

One of the earlier studies conducted on the relationship between short interest and returns was Asquith and Muelbroek (1995). They investigate the information content of short interest by seeing whether firms that are heavily shorted experience negative returns. They look at NYSE and AMEX stocks from 1976 to 1993 and find that this relationship holds. They find that short interest does indeed convey negative information and find that stock prices reflect positive information much better than negative information. In other words, stock prices tend to react much more to negative than positive information.

Desai et al. (2002) examine the relationship of short interest and stock returns on the Nasdaq US stock exchange between 1988 and 1994. They find that firms that are heavily shorted experience negative abnormal returns when accounting for market size, book to market and momentum factors. Desai et al. (2002) also find that firms that are heavily shorted are more prone to delisting compared to their respective matched size, book to market and momentum counterparts.

[25] As an example, if a particular stock had an open short interest of 30,000 shares and a total number of shares that could be short of 300,000, we would divide 300,000 by 30,000 and divide by 100 to obtain 10%. This means the Short Interest Ratio is 10% of float outstanding.

Adding on, Asquith et al. (2005) find that stocks that are experiencing a high short interest ratio and low institutional ownership (stocks which are owned more by retail investors over banks and hedge funds) underperform during 1988 to 2002 by 215 basis points per month on an equal weighted basis and 39 basis points per month on a value weighted basis. When value weighting is used the underperformance becomes virtually insignificant from significant.

Further on, Ackert and Athanassakos (2005) explore the relationship between short interest and stock returns in the Canadian stock market. Their results show that there is negative correlation between short interest and stock returns, i.e. as short interest increase, stock returns tend to decrease. Ackert and Athanassakos (2005) also find that returns for smaller firms are more negative as the supply of shares which are shortable are in less freely available on the open market. They also find that returns are less negative for those firms which options and convertible bonds on the open market. Overall their results suggest that market efficiency in regards to price discovery and price stability can be improved with less restrictions on the regulations on short sales.

Haruvy and Noussair (2006) find that by relaxing short selling constraints prices are lowered in asset markets, but prices are less reflective of fundamentals. They put forward the case that prices in asset markets are influenced by the restrictions on short selling and the limit of cash that is available to use to purchase.

Akbas et al. (2008) test the Miller (1977) Hypothesis against the Information Hypothesis to see what causes the negative abnormal returns that is usually associated with high levels of short interest. According to Miller (1977) stocks are overvalued in the presence of short-sale constraints and the negative abnormal returns are a correction of this overvaluation. On the other hand, the Information Hypothesis is based on the fact that short sellers are highly informed traders and short interest will therefore predict future returns due to the content of information. Akbas et al. (2008) find support for the later hypothesis.

Further on, Boehmer et al. (2008) show that heavily shorted stocks will on average underperform lightly shorted stocks by a risk adjusted[26] 1.16% over the following 20-day period. This works out to be 15.6% annualised. They see that the most informative short sales are the ones conducted by humans at institutions. Stocks which are heavily shorted by institutional holders will on average underperform by 1.43% the next month. Overall Boehmer et al. (2008) stress that short sellers are important contributors to efficient stock prices and the efficient market hypothesis.

Au et al. (2009) examine the relationship between short selling, returns and how arbitrage costs affect the behaviour of short sellers. Firstly, they use daily short selling data from the UK stock market to show that stocks that experience low levels of short interest experience positive abnormal returns on both an equal and value weighted basis. Au et al. (2009) state that short sellers tend to avoid stocks that have high firm specific[27] risk. They show that there exists a negative relationship between short interest and returns on stocks with

26 Risk-adjusted taking into account the return accountable for the risk being taken using a fair asset pricing model.

27 Firm specific risk is risk that is attached to a particular company over the general macro economy. A high level of debt may be firm specific but a tightening of monetary policy by a central bank would be considered macro specific risk, that could affect all firms regardless of debt level.

high firm specific risk and that short selling activity is more concentrated in stocks with low firm specific risk, where the cost of arbitrage is lower.

Diether et al. (2009) use tick data on short sales executed in the US stock market since 2005. They find that portfolios that are long lightly shorted stocks and short heavily shorted stocks have positive abnormal returns, though there would be a considerable amount of trading required to capture those returns as portfolio rebalancing would be required each and every month to account for new short interest information.

Subsequently, Boehmer et al. (2010) find that stocks with relatively high short interest experience negative abnormal returns. In contrast they also find that heavily traded stocks with low short interest experience positive abnormal returns. These positive returns in lightly shorted stocks are often larger than the negative returns in highly shorted stocks. Thus, being long a lightly shorted portfolio and being short a heavily shorted portfolio and then rebalancing the portfolio each month can produce excess returns beyond that of a fair asset pricing model.[28] Since the publication of Boehmer et al. (2010) hedge funds may have adopted this strategy as a means of excess return, so studies like mine are also interested in whether these strategies are still valid or have they been arbitraged out.

More recently, Hodgkinson et al. (2013) uses a high frequency UK dataset from 2003 to 2010 to explore short selling and stock returns. Their findings suggest that shorting indicates evidence of stocks which are overvalued; significantly negative abnormal returns follow after a heavily short stock. Hodgkinson et al. (2013) also notes that these results do not hold for shorting that is seen around the ex-dividend date of stocks. This is natural since the payment of the dividend, will cause the stock to drop by the respective value. Market participants are aware of the dividend and so will not see a stock that is down because of ex-dividend as negative information. Hodgkinson et al. (2013) also find their results holding for special periods such as the 2007-2009 global financial crisis.

Barinov and Wu (2014) propose a risk based firm type explanation on why stocks with high short interest ratios have lower future returns. They say that these firms have lower returns because they are used as a hedge against expected aggregate volatility risk. In other words, these firms are used to reduce volatility in a portfolio.

Rapach et al. (2016) show that short interest is the strongest known predictor of aggregate stock returns. Short interest can generate gains of over 300 basis points per annum for the mean-variance investor. A 300-basis point return is equivalent to 3% extra return while taking on no further risk. A vector autoregression decomposition is also performed and shows that the predictive power of short interest stems from the cash flows.

Chan et al. (2017) use the Chinese markets during earning announcement periods to test the Miller (1997) hypothesis that stocks are seen to be overvalued in the presence of short sale constraints. They find that stocks with short sale constraints have negative abnormal returns. Their findings help to explain the effect of short sale constraints on pricing efficiency.

28 In this case Boehmer et al. (2010) used the Fama French Three Factor Model with Momentum as their fair asset pricing model

Finally, Hendershott et al. (2018) show that short selling corporate bonds forecasts future bond returns. In particular this is true for high yield bonds, where private information is more likely and may not have been incorporated into the bond price. These results show that bond short sellers are contributing to efficient bond prices, however bond short selling does not predict returns for respective stocks but the converse holds true. This shows an information flow from stocks to bonds on future returns, but not from bonds to stocks.

We see that in the literature, short interest is used as an essential metric to compare short sellers and future returns. However, this is not the only metric, and often individual short sale trade data is also employed to gauge the effectiveness of short sellers. Using individual short sale trade data can be computationally taxing, as millions of trades are processed in a day per security and a subset of that volume being short sales is large. This means often smaller samples are used. Nonetheless individual short sale trade data is a great addition to the literature.

2.2.2 Short Selling and the Relationship with Liquidity, Price Discovery and Volatility

The short selling literature has also focused on liquidity, price discovery and volatility. Liquidity being how easily the security can be bought and sold. A deep order book on the buy and sell side in necessary to keep liquidity up, a deep order book is achieved by having limit orders [29] at varying prices and of sufficient quantities. Price discovery is the time taken to achieve a stable price by market participants, this is very much linked with liquidity. If price discovery is severely hampered, a security may struggle to open even for trading. Volatility is seen as the movement of a security over a finite period of time, often measured as the beta of the security (the movement of a security with respect to the market as a whole). The below papers outline some of the important work that has been done in this field of finance.

Initially, Diamond (1987) models the effects of short sale constraints on the change of stock prices to private information. Stopping traders from shorting reduces the effect of price adjustment to private information.

Ho (1996) explores short sales restrictions on volatility using the Singapore Stock Exchange as a case study for this experiment. The Singapore Stock Exchange suspended trading for three days in 1985, once the exchange had opened contracts could only be done on a basis of immediate delivery (meaning settlement within 24 hours). This meant that short selling was severely restricted for the duration of a month. Ho (1996) finds that while this short selling was restricted, the volatility of stock returns increased.

Bris et al. (2007) looked at 46 equity markets around the world and found that markets are much more efficient when short selling is allowed to take place. This paper is of particular interest to us as it is one of few papers that focuses on international evidence in regards to short selling bans, as the vast majority of papers have focused on the US markets. Bris et al. (2007) use a sample of 46 countries with time series as

[29] Limit orders are different from “at best” market orders. Limit orders are only executed when their price is hit by the security, “at best” market orders look for the best buy or sell price and execute immediately on the exchange given there exists a buy or sell price at all. Level 2 data from stock exchanges shows limit orders and thus the depth of the order book, Level 1 data shows the best buy and sell price in the order book.

well as cross sectional differences in short selling practises. The measurement of R squared and cross autocorrelations is used to calculate for asymmetry in stock returns, which is an indicator of short sales affecting price discovery. They find that short sale restrictions stop prices going further down. Their finding was that stock markets incorporated information much faster where short selling was allowed. Regulators such as the SEC and FCA often promote that short selling restrictions can reduce a sharp market decline, such as the decline seen in 2008 in the US equity market. Bris et al. (2007) find that market returns have less negative skew in markets where short selling is banned, i.e. there are more positive than negative returns when short selling is banned. Thus, markets are more efficient when investors are able to take short positions, contrary to the view that market regulators take.

Further on Chang et al. (2007) explore the effect of short sale restrictions on returns and volatility on the Hong-Kong market and found that the restrictions lead to higher volatility and less positive skewness of individual returns.

During the financial crisis of 2007-2009, Boehmer et al. (2008) show that short selling accounts for 13% of overall NYSE trading volume, and this percent holds across similar market capitalisation groups. Boehmer et al. (2009) show that this short selling volume was around 40% in 2007. Clearly short selling is very common among strong bear markets and lessens during bull markets. Short sellers are susceptible to what is known as a short squeeze, where open short positions are quickly closed out as the price of the security moves up. This can dramatically decrease short interest as the market from the bottom of a bear market into a new bull market.

In the same year, Clifton and Snape (2008) explore the effect of short selling restrictions on liquidity. They explore this issue via the Financial Services Authority short selling ban on financial and insurance stocks in 2008 in the UK during the financial crisis. They find that the average spread on banned stocks increases by 140% from 15 basis points to 36 basis points compared to a 56% increase for non-banned stocks from 14 basis points to 20 basis points. Depth of the order book deteriorated in both banned and unbanned stocks but more so for banned stocks, 59% vs 43%. Volume fell in banned stocks while it rose in unbanned stocks. Two separate regressions show that the decline in liquidity is independent of market changes such as increased volatility, therefore banned stocks observe less liquidity than their unbanned counterparts.

Lobanova et al. (2010) look at the impact of short sale restrictions on volatility, liquidity and market efficiency using the 2008 US short sale ban as evidence. Their results suggest there were statistically significant changes in liquidity, volume traded and return volatility while the short sale ban took place. They found market liquidity deteriorated substantially with spreads widening. Turnover ratio, volume traded and dollar volume declined during the short sale ban. Returns also decreased and volatility of those returns increased, therefore they suggest a significant drop in market efficiency has occurred.

Moving on, Battalio and Schultz (2011) use the 2008 US short sale ban to look at the impact on equity options markets. The stocks banned for short selling experience larger bid-ask spreads on their respective options. This is a good way to gauge liquidity of stocks as you are able to see the difference between the spread in options in banned and unbanned stocks. Options contracts will in turn mimic their underlying security. They also find that synthetic share prices for banned stocks become lower than actual share prices during the ban.

Adding on, Brockman and Hao (2011) look at the relationship between short selling and price discovery. Some underlying shares representing their respective American Depositary Receipts (ADRs) can be sold short in their home market, while others cannot. Brockman and Hao (2011) find that ADR short selling on American exchanges is more informative when those ADRs can't be sold short in their home market. This suggests that short sellers make a good contribution to price discovery.

Marsh and Payne (2012) add further light to the short sale ban of financial and insurance stocks in the UK equity markets. Again, like Clifton and Snape (2008), one of few studies conducted on the 2008 short sale ban on specifically UK equities. They find that order flows were not affected and that banned stocks were sold off even more aggressively compared to their non-banned counterparts once the short sale ban took place. Order book liquidity was also seen to have weakened and price discovery was hampered, these effects are seen to be reversed once the short sale ban was reversed. A weak order book is a core consideration of a stock exchange in regards to liquidity. Marsh and Payne (2012) hold a negative view towards the UK short sale ban and blame it for the deterioration in quality in UK equity markets for banned financial and insurance stocks.

In another investigation on price discovery, Boehmer and Wu (2013) use shorting flows to see the impact short sellers have on price discovery. Firstly, as shorting flow increases, intraday informational efficiency of prices improves. Second, at monthly and annual time frames more shorting flow increases the addition of public information into prices faster. Thirdly, greater shorting flow leads to lower post earnings announcement drift for earnings misses. Overall, they find that stock prices are more accurate when short selling is allowed to take place, therefore short selling according to Boehmer and Wu (2013) has a negative impact on price discovery in the market.

Beber and Pagano (2013) explore the effects of liquidity and price discovery during the 2008 short sale ban across the world and find that the short sale ban was bad for liquidity, in particular stocks with small capitalisation and those which have no respective options listed in the market. It also led to a slower price discovery process, especially in the case of bear markets and lead to the failure of price support, not including US financial stocks.

More recently, Callen and Fang (2015) use a large sample of US public firms and find robust evidence that short interest is related positively to one-year ahead stock price crash risk. Additionally, they find that this positive relation is more evident with excessive risk-taking behaviour, weak governance mechanisms and information asymmetry [30] between managers and shareholders. The high short interest is evident of market sentiment here, a security with high short interest is generally a riskier security to own.

[30] Information asymmetry is termed when one party has more information that the other, this can be evident when the managers are significant shareholders in an institution and may potentially hold back information to protect their shareholding interests. The market does however keep a key eye on insider selling and buying and this can often be a sign that a security is either under or overvalued, as managers are rarely likely to commit their own capital unless their stock is undervalued or at least at fair value.

Lee and Wang (2015) investigate the daily short selling by foreign investors and the impact they have on stock price, liquidity and volatility in the stock market of Korea. For the period of 2006 to 2010 they observe that the majority of short sales are performed by foreign rather than domestic market participants. Lee and Wang (2015) observe that short selling by foreign market participants occurs when the pressure to buy is high, however this does not help to improve stock liquidity. Volatility is not seen to increase by foreign short sellers, showing that foreign short sellers are not a destabilising presence in emerging markets.

In the same year, Nezafat et al. (2015) show that short sellers and active long-term investors (investors actively taking long positions in the market) are often in disagreement. They show that more than half of heavily short stocks have long positions by hedge funds from 2000 to 2011. Nezafat el al. (2015) show that stocks that are heavily shorted but have high hedge fund holdings do not underperform and the underperformance is only found in stocks that are heavily shorted but have low hedge fund holdings. These results show that active investors and short sellers often disagree.

Later on, Alves et al. (2016) measure the impact of the August 2011 financial stocks ban on covered short sells in the European countries of Belgium, France, Italy and Spain. The ban was put in place for 15 days. The short selling restriction did not reduce the volatility but that the volatility in those financial stocks increased instead, causing more problems for market regulators. Liquidity was also affected negatively in these stocks. The price discovery was affected as well as these stocks were shown to take longer to take in negative information that had entered the market.

Bohl et al. (2016) focus on the impact of short selling restrictions on stock return volatility. They apply two versions of their asymmetric markov-switching GARCH model [31] on financial stocks between 2008 and 2010 on the Frankfurt Stock Exchange in the German stock market. They find that although the financial crash lead to periods of increased volatility for the market as a whole, these stocks with short selling restrictions experienced more volatility than the market as a whole. Bohl et al. (2016) stress that short selling restrictions are more destabilising than they are stabilising, resulting in the market behaving more irrationally.

More recently, Sochi and Swidler (2018) look at the effects of price discovery using securities on the Dhaka Stock Exchange which have been in a constant short sale ban in comparison to securities on the New York Stock Exchange in which short selling is allowed. They employ a Monte Carlo Simulation on both sets of data using their respective indexes in regards of consecutive positive or negative returns (runs). They find fat tails in their run distribution where runs of longer duration appear on securities on the Dhaka Stock Exchange, showing that market efficiency may be impacted due to the short sale ban. This is because fat tails are indicative of a delay in information in pricing securities, which can occur from a short sale ban.

In addition, Zhisheng et al. (2018) explore the effects of the introduction of short selling which occurred in China in March 2010 for designated stocks. They explore the effects of short selling on price efficiency and liquidity. They find that the introduction of short selling improves price efficiency in terms of adjustment of

[31] GARCH modelling was proposed by Bollerslev (1986) and has had wide implications in modelling volatility using past asset returns. However recent studies have shown that estimates of GARCH models can be biased by structural breaks in volatility dynamics that typically occur during financial turmoil. Estimating GARCH on structural break data yields a non-stationary model and hurts predictive capabilities. A way to overcome this is with markov switching GARCH models whose parameters vary over time according to some regimes. These models can adapt to changes in unconditional volatility and improve predictions.

stocks to news. They also find short selling enhances liquidity based on bid-ask spread and Amihud illiquidity measure[32]. Overall short selling for them improves market quality even in less developed markets such as China.

The effect of short selling on liquidity, volatility and price discovery is often explored using ban periods. We see the general consensus of the literature has been short sale bans have negative effects on both liquidity and volatility. Short sale bans also affect the price discovery process. Some studies report volatility is increased in short sale bans, while others show no effect. Most studies report liquidity has deterioration in short sale bans.

2.2.3 Short Selling and Market Manipulation

The areas of concern here have been whether short sellers are informed and thus whether they engage in market manipulation. Being informed means short sellers are trading on available information, rather than manipulating a price by trying to simultaneously drive a stock lower. It is also important to look at the political aspects of short sale bans, the morality of short sellers and the view of the government in regards to short seller morality. This can affect whether short sale bans are implemented. The papers below outline these concerns.

Goldstein and Geumbel (2008) look at short sellers that target specific companies to depress their stock prices. They say the depressed prices in stock lead to distorted investment decision, therefore harming fundamentals and allowing short sellers to be able to cover their open position at further depressed prices. This is in a way similar to a negative feedback loop. Goldstein and Geumbel (2008) label this as a form of market manipulation and is indicative of certain but not all short sellers.

Comerton-Forde and Putnins (2009) examine how informed short sellers are, a proxy on market manipulation. They directly measure the abnormal returns earned on short sales on the NYSE. They find that short sellers are very good at predicting negative future returns and improve the efficiency of the market by counteracting overpricing. Commerton-Forde and Putnins (2009) also find that short sellers are more informed in stocks with smaller market caps, low book to market ratios and low analyst coverage. It is worth noting that they believe some naked short sellers[33] may be engaged in market manipulation. An overall take away from their study is that restricting short selling is likely to affect market efficiency but restricting naked short selling may help to increase the integrity of the market.

During the same year, Fotak et al. (2009) look at the level of fails to deliver during settlement as a proxy for naked short selling. They find that naked short sellers reduce volatility and price erroring and thus have positive effects on the quality of a market. They also look at the levels of naked short selling with four

32 Amihud illiquidity measure is a popular measure of illiquidity in a stock using returns data and daily volume. However, the bid-ask spread where available is preferred as it gives the true value of liquidity. In many securities it is difficult to obtain the bid and ask prices for stocks over a period of years and may be computationally difficult with the number of trades involved in a single day.

33 Naked short selling is termed when a short sale occurs without first borrowing the security from the broker. It is however not even necessary to borrow securities to sell short, since derivatives contracts such as call and put options mean long or short exposures can be held without owning or lending the security as a whole.

financial firms that were affected negatively during the 2007-2009 financial crisis. They find the levels of naked short selling were higher after the price declines and thus not responsible for the price declines. Fotak et al. (2009) maintain that not all naked short sellers are bad.

Adding on, Shkilko et al. (2009) look at the effects of short selling during intraday liquidity crises. They find that short sellers have negative destabilising effects on stock prices. Short sellers are seen to be amplifying intraday price decreases which are unrelated to information. Shkilko et al. (2009) suggest that what they have found is consistent with some short sellers being predatory.

In political aspects, Sirri (2009) considers various reasons that the SEC has changed from their initial position to balance efficiency and market quality issues and implement a short sale ban which strictly breaches market quality. In particular Sirri (2009) highlights the role of external sources such as vested political interests on the SEC which contributed to a strict short sale ban, when it may have been otherwise not necessary to implement a ban. They note that with the influence of the US Congress on the SEC, lobbying interests from short sellers would be targeted at the US Congress. It is often seen that the independence of the SEC and US Federal Reserve is tied to the US Treasury, which is under the control of the US Congress.

Subsequently, Engelberg et al. (2012) finds that short sellers are very adept at analysing public information and this is the main component of their trading advantage. They use a database for short sales in combination with another database of news releases and find that the negative relation between future returns short sales is twice as large on news day and four times as large on days that contain negative news. Engelberg et al. (2012) notes that this public news provides good trading opportunities for short sellers who are seen to be informed traders rather than market manipulators.

Looking at the political aspect, Beber and Pagano (2013) note that short sale bans occurred during different times and for different lengths. This is often a politically driven event, as there is no set schedule when a short sale ban should start and when a short sale ban should end. It may have been good to follow the SEC in instigating a short sale ban but many nations choose bans of different length and time periods, with decisions taken by respective parliamentary authorities. For example, Spain conducted a short sale ban in the 2007-2009 financial crisis after the United States, while the United States lifted the short sale ban before Spain. Beber and Pagano (2013) note this difference and find that the short sale bans were detrimental for liquidity and price discovery. In particular stocks with small capitalisation and no listed options (call contracts and put contracts) available for trading were affected the most negatively.

In further regards to political aspects Howell (2016) looks at short selling restrictions in the EU and the US. Howell (2016) finds that the legislative process had major impacts on shaping the rules of short sale restrictions in both the EU and the US. In the US it can be seen with the reimplementation of a tick test (a test governing rules on when a short sale is allowed to take place), which was implemented by the SEC under the direction of the US Congress. In the EU it is seen in the short selling policies on sovereign debt in comparison to different policies on short selling with corporate bonds. Both debt instruments, but holding very different rules due to political aspects. Overall it is clear that politicians with little knowledge or experience in financial markets have a great say on rules regarding short selling.

More recently, Chan et al. (2017) examine informed vs speculative trading among short sellers prior to the downgrade[34] of analysts in china for the observation period of March 2010 to August 2014. Given their full sample they do not find uninformed short selling in the days before the downgrade announcements. They split their findings into a star and non-star analyst sample and find that in the star analyst sample there is uninformed short selling, while for the non-star analyst sample there is not. They also observe that short positions are cut when star analysts upgrade a stock incorrectly, so it seems short sellers are not any more informed than the star analysts. Their evidence suggests that star analysts may be leaking information to short sellers before the downgrades. Overall short sellers are seen to be informed by Chan et al. (2017).

2.3 Conclusion

I conclude that my literature review above has been through and has shown the vast areas of finance I am studying in order to produce this thesis. My study differs from earlier work, to my knowledge it is the first study to see whether the Fama and French Five Factor Model shows whether adopting the long/short monthly rebalanced short interest portfolio strategy of Boehmer et al. (2010) is conclusive of excess return beyond which is specified by this fair asset pricing model. My study also is the first to judge whether the strategy of Boehmer et al. (2010) has been arbitraged away by market participants or whether the abnormal excess returns remain.

My study will be one of few to use international evidence in support, as most studies on short interest are focused on the American markets.[35] I will be one of few studies studying the effects of a short sale ban on liquidity, volatility and price discovery, using unique specified methods for each metric.

We have seen how the asset pricing literature has built up from the CAPM and extended onto the APT theory. We have seen how Fama and French (1993) use the foundations of the CAPM and incorporate factors of price to book ratio and company size to extend the CAPM to focus on where returns are truly being generated. We see how Carhart (1997) extends this further with the momentum factor to incorporate areas of behavioural finance. Finally, we see Fama and French (2015) further extend the Fama and French Three Factor Model to included profitability and investment.

On the short selling literature side, we see the work of Desai et al. (2002), Asquith et al. (2005) and Boehmer et al. (2010) using short interest to gauge future stock returns, Hendershott et al. (2018) employs the same technique across the corporate bond market. We see how high short interest stocks and bonds underperform low short interest stocks and bonds across many studies.

---

[34] An analyst downgrade is when a financial analyst believes that a stock will perform less than they expected from the last time it was reviewed. Therefore, a lower target price the analyst believes the stock will achieve in a year time is set. Downgrades can be given a rating of either Hold or Sell, where hold means to hold a stock if bought and refraining from buying if not bought. Sell means to sell a stock if bought and refraining from buying if not bought.

[35] The focus in academic finance literature has always been towards the American markets once the United States became a dominant superpower after World War 2 (WW2). The United States has the largest financial market in the world in terms of volume traded and market capitalisation across the Amex, Nasdaq and NYSE. The market capitalisation of the American markets is over 30 trillion dollars as of 2019.

We see how short sale bans are used to gauge the effects of short selling on liquidity, price discovery and volatility. Ho (1996), Chang et al. (2007) and Bohl et al. (2016) focus on volatility, while Clifton and Snape (2008), Battalio and Schultz (2011), Marsh and Payne (2012) and Beber and Pagano (2013) focus on liquidity. Brockman and Hao (2011), Boehmer and Wu (2013) and Sochi and Swidler (2018) focus on price discovery. Lobanova et al. (2010) focus on all three metrics. This topic has been explored in great depth across many markets and with many researchers. Generally, it is well documented that short selling bans have detrimental effects on price discovery, volatility and liquidity, where the best course of action has been to not implement a short sale ban at all and allow the market to freely trade.

We see how short sellers have often been labelled in the past as market manipulators and by seeing whether they are trading on good public information, may alleviate these misconceptions. Comerton-Forde and Putnins (2009) and Chan et al, (2017) with many other studies focus on the informed nature of short sellers. We see that although the literature generally sees short sellers are informed, there are exceptions and there may be a degree of market manipulation by a small minority of short sellers. Politically short sellers have come to the attention of governments and their vested interests are affecting market regulation. In particular, Sirri (2009, Beber and Pagano (2013) and Howell (2016) note these political concerns with short sale bans.

This section concludes the review of literature and helps to build my theoretical framework. The topics covered in this literature review are based on my four research questions, though there is a great quantity of overlap between topics that needs to be considered. Overall the literature states the importance of asset pricing models in determining a fair rate of return, it also states the importance of short sellers, the negative effects of short selling bans and the pollicization of the process.

CHAPTER 3: THEORETICAL FRAMEWORK

The theoretical framework chapter is concerned with more the theory behind the finance than a review of the literature. It outlines the models and methods I am using and the mathematics behind them. It also includes the theory of other areas of finance that are relevant to this study. A literature review is very good at giving background information on the current state of play in the literature, but will seldom explain the theories involved, this is what my theoretical framework is aiming to achieve.

3.1 Models Employed

This section is most concerned with the models that are being used in the study. The reader will be able to appreciate the mathematics behind such models and how they were derived given the state of play then in the finance literature. It will also allow the reader to see the drawbacks and limitations of the models I am using and the assumptions that many models depend on. The main models employed are to do with asset pricing models in the discrete time frame. We will also see models that are very much dependent on econometric theory such as GARCH as well.

3.1.1 Capital Asset Pricing Model

The majority of the current models in the asset pricing literature build upon the Capital Asset Pricing Model (CAPM), therefore it is very important to see the theoretical framework behind such a model. The CAPM was developed as a tool to work out the expected return of a security, given the risk of a security in the market, i.e. the fair return for the level of risk taken.

The CAPM derives the general equilibrium solution under the assumptions that investors are mean variance optimisers and that markets are perfect. Mean variance optimisers meaning investors are interested in maximising returns and minimising variance and that markets hold set conditions to make them perfect such as rational expectations, which I will build on. The CAPM holds that risk can be split as follows:

Total risk = Systematic risk + Unsystematic risk (1)

The CAPM says that unsystematic risk can be diversified away from holding an efficient portfolio, since everyone is able to diversify [36] and the cost of diversification is zero, the market should not be expected to reward someone for any unsystematic risk that is held. However, on the other hand, systematic risk cannot

[36] There is a saying in finance that there is "no free lunch", you should not be rewarded for a risk-free activity since the nature of the activity means everyone can do it, so the benefits of the activity will be arbitraged out by market participants. This in particular holds when markets are seen to be strong-form efficient according to the Efficient Market Hypothesis (EMH).

be diversified without taking on extra cost, therefore the systematic risk holder should be compensated for the extra systematic risk they are taking on. Therefore, according to these principles, the equation for the CAPM is stated as follows:

$$E(R_i) = R_f + \frac{COV(R_i,R_m)}{Var(R_m)} \times \left[E(R_m) - R_f \right] \tag{2}$$

Where $E(R_i)$ is the expected return on asset *i*. R_f is the equilibrium risk free rate of return. $\frac{COV(R_i,R_m)}{Var(R_m)}$ is the systematic risk of asset *i*, sometimes known as just the risk of asset *i*. The expression $\left[E(R_m) - R_f \right]$ is the market price of risk, that is the excess return above the risk-free rate that holding the entire market portfolio is able to achieve. $\frac{COV(R_i,R_m)}{Var(R_m)}$ measures the contribution to the covariance of the portfolio with the market portfolio as a fraction of the total variance of the market portfolio (Bode et al, 2015).

The model is often written more simply as:

$$E(R_i) = R_f + \beta_i \times \left[E(R_m) - R_f \right] \tag{3}$$

where:

$$\beta_i = \frac{COV(R_i,R_m)}{Var(R_m)} \tag{4}$$

β_i is referred to as beta and represents the systematic risk of asset *i*. So essentially the CAPM states a fair rate of return is the return of the risk-free rate plus the excess market return multiplied by the volatility of the asset with respect to the market. So according to the CAPM model there are theoretically only several ways of gaining extra return. The first being a higher risk-free rate, which is often determined by the base rate in central banks (the rate at which the central bank lends to recognised commercial banks). The second being a higher excess market return, which is determined by the demand for holding the market portfolio as opposed to holding other fixed income securities or commodities as a whole. The third being holding an asset with a higher beta and thus the asset has historically been more volatile than the market as a whole.

Mathematically beta is calculated by taking the covariance of the security with respect to the market and dividing by the variance of the market. It is worth noting that beta is not consistent over time in the sense that the beta of a security is constantly changing. Beta makes use of historic data [37] and there is no guarantee

[37] Historic data is often used for financial metrics such as beta, however the past is not exactly the best indicator of the future, so forward-looking metrics have also been developed such as Forward P/E which works on projected earnings rather than past earnings. Projected earnings are however an estimate and can change drastically over time due to firm specific and macro uncertainties. Markets are seen to be always forward looking, so these metrics are of great use.

a security will maintain a beta that it once had. This is often the case in stocks that experience decades of growth and reach a point where the size of the company limits growth, and thus the beta of the stock often falls with the slower growth. The level of speculators in a stock can also drive its beta, and thus stocks with lower amounts of speculators hold a significantly lower beta.

The CAPM as a model is built on a series of assumptions that must be held for the model to remain valid. These assumptions relate to investors and a perfect market. Regarding a perfect market, the market must be frictionless and hold perfect information. No information must be withheld from market participants and information when it comes available must reach all market participants at the same time. There must also be no imperfections like regulations, tax or indeed restrictions on short selling. Restricting short selling impacts the efficiency of the markets and thus impacts the CAPM. Another assumption regarding a perfect market is that all assets are publicly traded and perfectly divisible. The final assumption is that there is perfect competition and that everyone is a price taker. If the security is at an undervalued price, there will be a sufficient supply of buyers, if a security is at an overvalued price, there will be a sufficient supply of sellers.

There are also a set of assumptions regarding investors. Investors should have a same one period horizon in their investment ideology. They should be rational and be wanting to maximise expected utility over a mean variance space. They must also finally hold homogenous beliefs (Bode et al, 2015).

The CAPM is the backbone of most asset pricing models. Most asset pricing models agree on the market risk premium factor $\beta_i \times [\,E(R_m) - R_f]$, however a lot explain that returns are driven by more than just that coefficient. For a long time, the CAPM remained the standard for asset pricing models, however now it has been superseded by larger factor models and continuous time finance. Back testing is a technique that has often been used to derive what drives returns, and the suitability of models are compared with back testing.

Over recent times the CAPM has often been seen to be redundant and that it does not reflect fair asset pricing returns even though many academics see it as a good benchmark. The relationship between beta and market returns is often criticised for being often too flat or even negative, which in the standard CAPM model should not hold. With a flat beta, an investor is compensated little for taking on extra risk, therefore putting forward the case that investors may prefer less risky assets given the lack of benefit of riskier assets. A negative beta would ensure that investors prefer less risky assets and are ill advised to take on riskier assets. This phenomenon of a flat beta may be due to the fact that equity markets are highly correlated with negative macro news often impacting every stock in the market. Selling pressure on one area of the market can often spread to other areas. One of the most common reasons for selling pressure is because everyone else is selling and the only way to limit losses is to also sell.

3.1.2 Fama and French Three Factor Model with Momentum

The Fama and French Three Factor Model with Momentum was the work of Eugene Fama and Kenneth French extended by Mark Carhart to form an asset pricing model that can better model historic returns than the CAPM. The Fama and French three factor model is often written as follows:

$$E(R_i) - R_f = \alpha_i + \beta_i\left[E(R_m) - R_f\right] + \gamma_i E(SMB) + \delta_i E(HML) + \varepsilon_i \tag{5}$$

Where $E(R_i)$ is the expected return on asset *i*. R_f is the equilibrium risk free rate of return. β_i is the systematic risk of asset *i*, sometimes known as just the risk of asset *i*. $\left[E(R_m) - R_f\right]$ is the market price of risk, the excess return above the risk-free rate that holding the entire market portfolio is able to achieve. γ_i is the factor loading for SMB (Small Minus Big). SMB being the factor that notes the difference between the outperformance of small firms compared to large firms. δ_i is the factor loading for HML (High Minus Low). HML being the factor that notes the difference between the outperformance of growth stocks compared to value stocks. ε_i is the error term to account for inconsistencies in the model (Fama and French, 1993).

The Fama and French Three Factor Model says that the expected return for a security is a combination of the risk-free rate plus the excess return of the market component plus the small minus big component plus the high minus low component. As we can see this model is a direct extension from the CAPM, with the second term on the right-hand side of the model and the terms on the left-hand side of the model being the terms from the original CAPM.

The Fama and French Three Factor Model suggests that the SMB and HML are the factors that contribute to the systematic risk that is not captured by the CAPM. If the model fully explained returns it was able to hold, then there would be no requirement for an alpha term. The alpha term is often included in the model for the return that can't be explained by the model that is not attributed to the error term. α_i would be equal to zero if the model fully explained returns.

The factor loading SMB and HML are created from portfolios. Firms are sorted by size and the book to market ratio. SMB is constructed as the difference between the returns of small firms vs large firms, while HML is constructed as the return between high and low book to market firms. The excess return on market factor is taken as the return of the entire market (NYSE, Amex and Nasdaq) minus the treasury bill rate.

The Fama and French Three Factor Model is constructed using six value weight portfolios formed on size and book to market. The construction is very similar to the Fama and French Five Factor Model, although much simpler since fewer factors are being used in the construction of the model.

Table 1: Portfolios formed by Size and Book to Market Ratio

Book to Market Ratio	Size	
	Small Firm	Big Firm
Value	S/V	B/V
Neutral	S/N	B/N
Growth	S/G	B/G

Notes: Data Source from French (2017), Kenneth French Library. Table 1 shows the six portfolios formed by using size and book to market ratio. The six portfolios are used in the construction of the SMB factor. In regards to book to market ratio, *V* stands for Value, *N* stands for Neutral, *G* Stands for Growth. In regards to firm size, *S* stands for Small and *B* stands for Big.

We can see the portfolios formed by size and book to market ratio above in **Table 1**.

The SMB factor is formed from the average return on three small portfolios minus the average return on three big portfolios, mathematically it is constructed as follows:

SMB = 1/3 (Small Value + Small Neutral + Small Growth) – 1/3 (Big Value + Big Neutral + Big Growth)

The HML factor is formed from the average return on two value portfolios minus the average return on two growth portfolios, mathematically it is constructed as follows:

HML = 1/2 (Small Value + Big Value) – 1/2 (Small Growth + Big Growth)

This outlines the construction of the Fama and French Three Factor Model.

Table 2: Portfolios formed by Size and Book to Market Ratio in Testing

Book to Market Ratio	Size		
	Small Firm	Medium Firm	Big Firm
Value	S/V	M/V	B/V
Neutral	S/N	M/N	B/N
Growth	S/G	M/G	B/G

Notes: Data Source from Bode et al. (2015). Table 2 shows the nine portfolios formed by using size and book to market ratio. The nine portfolios are used in the first pass regression of Fama and French. In regards to book to market ratio, *V* stands for Value, *N* stands for Neutral, *G* Stands for Growth. In regards to firm size, *S* stands for Small, *M* stands for Medium and *B* stands for Big.

Next, I look at how the model is tested and what results come out from the testing. The Fama and French Three Factor Model is tested by forming nine portfolios which have sensitivities to each factor. A matrix of nine portfolios is formed by firms sorted into small, medium and big on the SMB side and value, neutral and growth on the HML side. We can see the portfolios formed by size and book to market ratio in testing in **Table 2**.

For these nine portfolios a first pass regression is run over 816 months between 1929 and 1977 on the model:

$$R_i - R_f = \alpha_i + \beta_i\left(R_m - R_f\right) + \gamma_i SMB + \delta_i HML + \varepsilon_i \tag{6}$$

Table 3: Performance of Individual Portfolios against Fama French Three Factor Model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Portfolio	B/M	Size	Ex Return	Alpha	Beta	Gamma	Delta	t(Alpha)	t(Beta)	t(Gamma)	t(Delta)	R^2
S/V	0.55	22.39	0.61	-0.42	1.06	1.39	0.09	-4.34	30.78	19.23	1.73	0.91
S/N	1.11	22.15	1.05	-0.01	0.97	1.16	0.37	-0.18	53.55	19.49	9.96	0.96
S/G	2.83	19.05	1.24	-0.03	1.03	1.12	0.77	-0.73	67.32	39.21	26.97	0.98
M/V	0.53	55.85	0.7	-0.06	1.04	0.59	-0.12	-1.29	55.83	18.01	-4.3	0.96
M/N	1.07	55.06	0.95	-0.01	1.05	0.47	0.34	-0.15	32.98	17.5	9.5	0.96
M/G	2.18	53.21	1.13	-0.04	1.08	0.53	0.73	-0.9	47.85	8.99	11.12	0.97
B/V	0.43	94.65	0.58	0.02	1.02	-0.1	-0.23	0.88	148.09	-6.88	-13.52	0.98
B/N	1.04	92.06	0.72	-0.09	1.01	-0.14	0.34	-1.76	61.61	-4.96	13.66	0.95
B/G	1.87	89.53	1	-0.09	1.06	-0.07	0.84	-1.4	52.12	-0.86	21.02	0.93

Notes: Data Source from Bode, Kane and Marcus (2015), Investments. Table 3 shows the performance of individual portfolios categorised by book to market ratio against size. In regards to book to market ratio, *V* stands for Value, *N* stands for Neutral, *G* Stands for Growth. In regards to firm size, *S* stands for Small, *M* stands for Medium and *B* stands for Big.

The results obtained from the first pass regression are shown as follows in **Table 3**. From the results we can see in Column (4) that except for the S/V portfolio, the alpha of all the portfolios is very small. They are also seen to be insignificant as their t statistic shown in Column (8) is below 2. We also have large R^2 values in Column (12) for the portfolios that shows the three-factor model does a very good job at explaining returns. We also see large t statistics on SMB and HML loadings in Columns (10) and (11) that show that these factors in particular contribute significantly to explanatory power.

The S/V portfolio is worth mentioning here as it has a significantly large alpha, showing a greater deal of unexplained return. Whether this is unique to the portfolio or a wider issue means more testing would be required around that portfolio to see the root cause of this. Overall, the Fama and French Three Factor Model does a good job of modelling past returns, and in particular a much better job than the CAPM.

Boehmer et al. (2010) used the Fama and French Three Factor Model with Momentum as their adjuster for risk premium, which till the recent introduction of the Five Factor Model has been the standard in asset pricing models. The Fama and French Three Factor Model with Momentum is written as follows:

$$E(R_i) - R_f = \alpha_i + \beta_i\left[E(R_m) - R_f\right] + \gamma_i E(SMB) + \delta_i E(HML) + \eta_i E(MOM) + \varepsilon_i \tag{7}$$

As we can see the model is very similar to the Fama and French Three Factor Model, with the only addition being the $E(MOM)$ term, i.e. the momentum term. Jegadeesh and Titman (1993) found that the performance of good and bad stocks persisted over a few months, an example of a momentum tendency. Carhart (1997) added this fourth factor to the Fama and French Three Factor Model and formed what is now known as the Fama and French Three Factor Model with Momentum or the Carhart Four Factor Model. Carhart (1997) found that the alpha of many mutual funds could be attributed to this momentum factor.

More recently Drechsler and Drechsler (2016) extend the Carhart Four Factor Model further to include a CME factor. CME stands for Cheap Minus Expensive and represents the return premium high short-fee stocks have had over low short-fee stocks, even net of fees. Therefore, their model is written as follows:

$$E(R_i) - R_f = \alpha_i + \beta_i\left[E(R_m) - R_f\right] + \gamma_i E(SMB) + \delta_i E(HML) + \eta_i E(MOM) + \mu_i E(CME) + \varepsilon_i \tag{8}$$

Drechsler and Drechsler (2016) believe that the CME factor is a great contributor to describe stock returns and does a better job than the standard Carhart Four Factor Model. We can see that models are being built with more and more factors, showing the complex nature of returns, during the same timeframe Fama and French (2015) is published building on earlier work.

3.1.3 Fama and French Five Factor Model

However, as it is seen with many financial models, the academic finance community was not content with the Carhart Four Factor Model and looked at even more datasets across many countries to see what other factors could contribute to stock returns. Fama and French (1993) was a great step up from the extensions of the CAPM, as it was one of the first models to truly lay down multiple factors responsible for stock returns in particular factors outside of beta. However, a degree of alpha or unexplained returns remained that their previous models failed to address.

Novy-Marx (2013) and Titman, Wei, Xie (2004) find that the Fama French Three Factor is not a complete model and that the Carhart Four Factor Model, although an improvement, fails to consider the returns related to profitability and investment. This was the primary motivation for Fama and French (2015) and the extension of their three-factor model into a newer and more recent five factor model. The model is given as below:

$$E(R_i) - R_f = \alpha_i + \beta_i\left[E(R_m) - R_f\right] + \gamma_i E(SMB) + \delta_i E(HML) + \eta_i E(RMW) + \lambda_i E(CMA) + \varepsilon_i \quad (9)$$

The base of this model builds up from the Fama and French (1993) Three Factor Model with two additional factors that of profitability (RMW) and investment (CMA). So, the two additional terms in the model are $E(RMW)$ and $E(CMA)$. RMW is the difference between the returns on diversified portfolios of stocks with robust and weak profitability. CMA is the difference between the returns on diversified portfolios of stocks of low and high investment firms. Hence Fama and French (2015) use RMW for robust minus weak and CMA for conservative minus aggressive.

Table 4: Portfolios formed by Size and Book to Market Ratio

Book to Market Ratio	Size	
	Small Firm	Big Firm
Value	S/V	B/V
Neutral	S/N	B/N
Growth	S/G	B/G

Notes: Data Source from French (2017), Kenneth French Library. Table 4 shows the six portfolios formed by using size and book to market ratio. The six portfolios are used in the construction of the SMB factor. In regards to book to market ratio, *V* stands for Value, *N* stands for Neutral, *G* Stands for Growth. In regards to firm size, *S* stands for Small and *B* stands for Big.

Table 5: Portfolios formed by Size and Investment

Investment	Size	
	Small Firm	Big Firm
Conservative	S/C	B/C
Neutral	S/N	B/N
Aggressive	S/A	B/A

Notes: Data Source from French (2017), Kenneth French Library. Table 5 shows the six portfolios formed by using size and investment. The six portfolios are used in the construction of the CMA factor. In regards to investment, *C* stands for Conservative, *N* stands for Neutral, *A* stands for Aggressive. In regards to firm size, *S* stands for Small and *B* stands for Big.

Table 6: Portfolios formed by Size and Profitability

Profitability	Size	
	Small Firm	Big Firm
Weak	S/W	B/W
Neutral	S/N	B/N
Robust	S/R	B/R

Notes: Data Source from French (2017), Kenneth French Library. Table 6 shows the six portfolios formed by using size and profitability. The six portfolios are used in the construction of the RMW factor. In regards to profitability, *W* stands for Weak, *N* stands for Neutral and *R* stands for Robust. In regards to firm size, *S* stands for Small and *B* stands for Big.

Fama and French Five Factor Model is constructed using six value weight portfolios that are formed on size and book to market, six value weight portfolios formed on size and investment and six value weight portfolios formed on size and profitability. **Table 4** above shows these six value weight portfolios formed on size and book to market, **Table 5** above shows these six value weight portfolios formed on size and investment and **Table 6** above shows these six value weighted portfolios formed on size and profitability.

The factors for the Fama and French Five Factor Models are formed from these portfolios. The SMB (Small Minus Big) factor is the average return of the nine small stock portfolios (S/V to S/R) minus the average return of the nine big stock portfolios (B/V to B/R). We can mathematically write down the construction of this factor in the following way:

$SMB_{B/M}$ = 1/3 (Small Value + Small Neutral + Small Growth) – 1/3 (Big Value + Big Neutral + Big Growth)

SMB_{OP} = 1/3 (Small Robust + Small Neutral + Small Weak) – 1/3 (Big Robust + Big Neutral + Big Weak)

SMB_{INV} = 1/3 (Small Conservative + Small Neutral + Small Aggressive) – 1/3 (Big Conservative + Big Neutral + Big Aggressive)

$SMB = 1/3\ (SMB_{B/M} + SMB_{OP} + SMB_{INV})$

The HML factor is the average return of two value portfolios minus the average return on two growth portfolios. It is formed in the following way:

HML = 1/2 (Small Value + Big Value) – 1/2 (Small Growth + Big Growth)

The RMW factor is the average return on the two robust profitability portfolios minus the average return on the two weak profitability portfolios:

RMW = 1/2 (Small Robust + Big Robust) – 1/2 (Small Weak + Big Weak)

The CMA factor is the average return on two conservative investment portfolios minus the average return on two aggressive investment portfolios.

CMA = 1/2 (Small Conservative+ Big Conservative) – 1/2 (Small Aggressive+ Big Aggressive)

This concludes the construction of the Fama and French Five Factor Model, it can be seen that although two extra factors have been added, it has substantially increased the complexity of the construction of the model due to permutations and combinations. Taking models even larger to 7 or 8 factors would exponentially increase the complexity and thus potential for error as each term can contribute to the error. We also have the issue with more terms that we may overfit a model and thus contribute to an R Squared Value that does not represent the true fit of a model. On an econometric basis this is worth keeping in mind. We have seen how vastly different the Fama French Models are constructed in comparison to the CAPM.

3.1.4 GARCH (p,q) Model

Generalised Autoregressive Conditional Heteroskedasticity (GARCH) models are often used in finance to model volatility. Volatility being the day to day fluctuations in security prices driven by the constant stream of buyers and sellers in any given market. Financial market security prices often exhibit a phenomenon known as volatility clustering, where time series data of security prices often shows periods of high volatility followed by periods of low volatility. Volatility clustering was first noted in Mandelbrot (1963) where he noticed that large changes seem to be followed by large changes, of either sign, and small changes tend to be followed by small changes of either sign. This is often true in periods of financial distress, where large sustained volumes of panic selling can cause securities to swing wildly. Once markets have calmed, volumes often drop and volatility is low. Volatility is often measured with the Chicago Board Options Exchange Volatility Index (CBOE VIX). The CBOE VIX shows the implied volatility of the market calculated using options contracts and the risk-free rate, the CBOE VIX shows a great deal of volatility clustering. Therefore, with the time varying nature of volatility rather than a constant stream of volatility, GARCH models in theory are ideal to model financial markets.

By using an autoregressive moving average (ARMA) model for the error variance in our model, it is possible to form a GARCH model. Therefore a GARCH (p,q) model consists of p order of GARCH terms and q order of ARCH terms. The GARCH terms lags are represented by ${\sigma_{t-i}}^2$ and the ARCH term lags are represented by ${\varepsilon_{t-i}}^2$ in the following model:

$$y_t = x_t'b + \varepsilon_t \tag{10}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \tag{11}$$

$$\varepsilon_t | \psi_{t-1} \sim \mathcal{N}(0, \sigma_t^2) \tag{12}$$

Where equation (10) is the mean equation and equation (11) is the conditional volatility equation. Further equation (12) states that errors of the GARCH model are distributed with a mean of zero and a variance that is modelled with an ARMA process that consists of both an autoregressive (AR) element and a moving average (MA) element. To test for heteroskedasticity the White Test is advised.

Fitting an adequate model to the data is a main element of GARCH models as there are both *p* and *q* terms which can take an infinite number of lags. A researcher could propose GARCH as their model choice, but they would also need to propose the lag length of the GARCH terms and the lag length of the ARCH terms. Several methods can be employed to fit a GARCH model, but AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) are the most popular choice.

To fit using either the AIC or BIC, several models would be run with the given dataset and either the lowest level of AIC or lowest level of BIC would be taken as the optimal model choice. Model impact means models up to GARCH (3,3) are usually run, as higher order terms are seen to have little effect on datasets which are not of decades in length using daily data. Computation is also an issue and becomes much more difficult with larger GARCH models since MLE (Maximum Likelihood Estimation) is used to find optimal model fit. MLE is generally computationally difficult and larger lag lengths add to that complexity. The log likelihood in the MLE can also be flat meaning a model cannot be fit to the data. This is less of an issue with simpler GARCH models, but can become more problematic with specialist models such as EGARCH (Exponential GARCH) or QGARCH (Quadratic GARCH) etc. with higher order terms.

To choose between AIC or BIC is based on the needs of the model. AIC is used when the GARCH model derived is intended for forecasting and BIC is used when the model derived is intended for explanatory modelling. It has generally been seen that higher order GARCH models are not superior to lower order GARCH models and often a GARCH (1,1) can be more than adequate to model the volatility clustering present in the data.

The AIC is given with the following formula:

$$AIC = -2 \log L(\hat{\theta}) + 2k \tag{13}$$

Where k is the number of estimated parameters in the model and $L(\hat{\theta})$ is the maximum value of the likelihood function for that particular model. AIC rewards goodness of fit shown by the likelihood function of $L(\widehat{\theta})$, but also includes a penalty of k with increasing the parameters of a model. This penalty is to discourage overfitting of a model.

The BIC is given with the following formula:

$$BIC = -2\log L(\widehat{\theta}) + klog(n) \tag{14}$$

Where k is the number of estimated parameters in the model, $L(\widehat{\theta})$ is the maximum value of the likelihood function for that particular model and n is the sample observations. BIC also awards goodness of fit with the likelihood function of $L(\widehat{\theta})$. The main difference between AIC and BIC is that BIC punishes to a greater extent larger amounts of parameters in a model as opposed to AIC. AIC is not always better than BIC, but the general consensus among the academic econometric community is a preference for AIC over BIC when fitting a model. BIC is seen to result in more underfitting of a model and AIC is seen to result to more overfitting of a model. Both conditions of under and overfitting are ideally to be avoided.

One of the biggest drawbacks of GARCH models is that they are not robust to time aggregation, which means that if a series of a daily frequency can be modelled with a GARCH (1,1), the same series with a monthly frequency should be modelled with a different GARCH model.

Over the years GARCH models have developed vastly to include various conditions that may be useful to financial market modelling. The earlier generation of GARCH models have been very good at replicating volatility clustering, however only the magnitude of the shock and not the sign affects the conditional volatility. Therefore, the first generation of GARCH models can't model the fact that good news decreases volatility and bad news increases volatility. However, not all returns data is seen as asymmetric and this needs to be considered when modelling for that. The more recent class of GARCH models include the EGARCH model of Nelson (1991), AGARCH of Engle and Ng (1993) and the QGARCH of Sentana (1995).

A dummy variable can often be implemented into an GARCH model to look at directional changes in volatility, i.e. whether there is an increase or decrease in volatility. Dummy variables are useful to model changes between two time periods in one set of data and can be implemented to any model. In my case the dummy variable would take a value of 0 and 1 respectively for the two periods that are being investigated (the ban and unban periods respectively) and would be added on to the conditional equation. The remaining mean equation of the GARCH model would stay the same. Therefore the conditional equation of the GARCH (p,q) would gain the dummy variable and look as follows:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + D_{SSB} \tag{19}$$

D_{SSB} is the dummy variable and would account for the positive or negative directional change in volatility.

3.1.5 Bid-Ask Spread Model

We have seen how volatility can be modelled with an GARCH and how a dummy variable can be implemented to monitor directional change, however liquidity is a different concept to volatility. Whenever a security is traded on an exchange in the financial markets, a price is quoted on the Level 1 data, i.e. the price of a stock reported by news outlets and other financial organisations. Many people fail to realise that a security will not have one price, but in fact two and the price quoted in the Level 1 data is the price at which the last trade took place. There is no guarantee the next trade will take place at that price or even whether a trade will take place at all.

The two prices that a security usually quotes is often referred to as the bid price and the ask price. The Bid-Ask Spread is the difference between the two prices quoted, the ask price being the immediate sell price of a security and the bid price being the immediate buy price of a security. These prices are referred to as market orders and are executed on an exchange immediately. Orders can be also entered into the buy or sell side of the order book and are referred to as limit orders. Limit orders provide depth to both buy and sell side of the order book and give liquidity to a particular security, market orders on the other hand remove liquidity by taking orders out of the order book. Order books are very much automated with limit orders entering the order book and market orders removing potential bids and asks from the order book.

Often the order book can be used to determine support and resistance levels for a security as limit orders waiting to execute can be seen[38]. If a large buy order is sitting far up the order book on the buy side, it can be seen as a level of support for a security, similarly if a large sell order is sitting far up the order book on the sell side, it can be seen as a level or resistance for a security. These orders can be removed, so levels of resistance and support are constantly changing. Orders in the order book can be seen by the entire market and thus provide transparency in the efficient running of a market.

The Bid-Ask spread is one measure of liquidity that is used, if the Bid-Ask spread was zero we would have a frictionless asset. An assumption such as a frictionless market is often needed in finance models such as the Capital Asset Pricing Model (CAPM) or the Black Scholes Model. The Bid-Ask spread of a security would be calculated in the following way (given you had access to offer and bid prices, which can be an issue in some smaller emerging markets):

$$\frac{offer-bid}{offer} \times 100 = Percent\ spread \tag{20}$$

[38] A great number of professional traders will trade simply from the order book (level 2 data), with minimal reference to the stock price chart (level 1 data). Round whole numbers are usually good areas of resistance and support, prices put in to buy or sell securities are usually put in round whole numbers. The minimum division of price that can be quoted for a security is 1 penny for a stock listed on the London Stock Exchange and 1 cent for a stock listed on the New York Stock Exchange or NASDAQ, this decimalisation has meant round numbers are often set as resistance and support levels.

The smaller the bid-ask spread, the greater the liquidity in a security this shows. This is very important for day traders, who move in and out of securities on a daily basis, as they have to compensate for the bid-ask spread on top of other transaction fees and taxes in order to make a profit. If the bid-ask spread is very wide, it can discourage day traders and only encourage long term investors to buy or sell the security. Some companies prefer this and keep stock prices high in order to discourage day traders, a good example is the Class A shares of Berkshire Hathaway[39] which trade for over $300,000 a share in 2019. We can use an example to illustrate this concept of bid-ask spread.

Say for example GBP/USD current bid price is 1.2804 and the current offer price is 1.2808, the percent spread[40] would be the following:

$$\frac{1.2808-1.2804}{1.2808} \times 100 = 0.031\% \tag{21}$$

If we are able to compare bid-ask spreads between two periods, we can have a gauge of the change in liquidity. This is the theory behind the Bid-Ask Spread Model.

$$S_t = c + \beta_0 \alpha r_t^2 + \beta_1 \alpha v_{it} + \beta_2 e v_{it} + \beta_3 D_{SSB} + \varepsilon_t \tag{22}$$

The regression model in equation (22) is a Bid-Ask Spread Model that has been used in Lobanova et al. (2010) to measure liquidity. S_t is the average spread for stocks at time *t*, αr_t^2 is the daily average return of the sample squared at time *t*, αv_{it} is the daily average volume of the sample at time *t*, ev_{it} is the excess daily average volume of the sample at time *t*. The excess daily trading volume is calculated by subtracting the daily trading volume at time *t* from the average daily trading volume from the duration of the study. c is the constant of the model. D_{SSB} is the dummy variable which takes the value of 0 or 1 for the two time periods being compared. If the coefficient of D_{SSB} has a negative value and is significant it can be proved there is a decline in liquidity from the first period to the second, if the coefficient of D_{SSB} has a positive value and is significant it can be proved there is an increase in liquidity from the first period to the second. A coefficient of 0 and significance for D_{SSB} indicates there is no change in liquidity. Finally, ε_t is the error term of the model. This is the basis of the model proposed by Lobanova et al. (2010).

Often it is difficult to estimate S_t given that bid and ask data can be hard to obtain for securities in emerging markets and even for many securities in developed markets. Therefore, it is best to use an estimator as a

[39] Berkshire Hathaway is a US multinational conglomerate that trades on the NYSE, with full ownership of businesses such as GEICO, Duracell, Dairy Queen and Fruit of the Loom. It also has large stakes in businesses such as Apple, American Express, Kraft Heinz and Wells Fargo. It is a component of the S&P 100, one of the largest 100 public firms trading on US exchanges.

[40] The spread in forex is often quoted in pips instead of percent, the spread of the GBP/USD pair above would be recognised as 4 pips. Since most currency pairs are quoted to 4 decimal places, 1 pip is used to represent a change in the fourth decimal point or 1/100th of 1%.

proxy for S_t . Corwin and Schultz (2012) propose a method to estimate the bid-ask spread using daily high and low prices for securities. The Corwin and Schultz (2012) bid ask spread estimator is as follows:

$$S = \frac{2(e^{\alpha}-1)}{(1+e^{\alpha})} \tag{23}$$

Where:

$$\alpha = \frac{\sqrt{2\beta}-\sqrt{\beta}}{3-2\sqrt{2}} - \sqrt{\frac{\gamma}{3-2\sqrt{2}}} \tag{24}$$

$$\beta = E\left\{\sum_{j=0}^{1}(ln\frac{H_{t+j}^{0}}{L_{t+j}^{0}})^2\right\} \tag{25}$$

$$\gamma = E\left\{\sum_{j=0}^{1}(ln\frac{H_{t,t+1}^{0}}{L_{t,t+1}^{0}})^2\right\} \tag{26}$$

By having this estimator, we can use high and low prices to work out a good approximation to spreads. High and low prices are readily available for most markets and this is a great benefit to our estimation methods. From the equations above, α is the high-low price estimator, β is the summation of the squared high-low price spread. γ is the squared high-low price spread. S is the stock's high-low bid-ask spread estimator. H_t is stock's high price on day t, L_t is the stock's low price on day t. $H_{t,t+1}$ is the stock's highest price over the two-day period of t and t+1. $L_{t,t+1}$ is the stock's lowest price over the two day period of t and t+1 (Corwin and Schultz, 2012).

Corwin and Schultz (2012) say that their estimator is far more accurate than the Roll (1984) covariance spread estimator and is very accurate under ideal conditions. When there are significant overnight returns and prices seem to be observed sporadically, the Corwin and Schultz (2012) estimator tends to underestimate spreads. Throughout realistic conditions Corwin and Schultz (2012) find that the correlation between their estimator and the true spread based on market data is 0.9. They also find that the standard deviation of their spread estimator is one half of the standard deviation of the Roll (1984) covariance spread estimator.

Ripamonti (2016) tests the Corwin and Schultz (2012) in the Brazilian Stock Market. Ripamonti (2016) shows that this estimator is a reliable gauge to estimate spreads. The findings show that the spread has stationary properties, allowing the ability to forecast in periods of lagged variables and also time-varying cointegration with book to market, debt to equity, size and return.

The Corwin and Schultz (2012) estimator is more suitable than the Roll (1984) covariance spread estimator when it comes to estimating spreads. In addition, its construction is relatively easy and can be implemented to most datasets with ease given the availability of high and low prices. A full derivation of the estimator, which is of length, can be found in Corwin and Schultz (2012).

3.2 Methods Employed

This section aims to outline the theoretical methods that have been employed for the estimation procedures. We have seen the models that will be used for my estimation procedures, but sound econometric methods need to be applied to these models to get results of value. The two main econometric methods employed in my research is of Ordinary Least Squares Regression (OLS) and Maximum Likelihood Estimation (MLE). The aim of both methods is to find the true estimator for a given set of data to the degree that this estimator can be used for interpolation and extrapolation.

3.2.1 Ordinary Least Squares Regression (OLS)

OLS is a technique used in regression analysis to find the best line of fit with a single independent variable and either a single or multiple dependent variable[41]. It is the simplest of regression techniques and is the foundation for more advanced regression techniques.

The regression techniques I employ involve multiple regressors, so I will outline the derivation for multiple regressors regarding OLS. In the case of multiple regressors the coefficient β becomes a vector. The Sum of Squared Errors $SSE(\beta)$ is taken as a function of the vector β. The aim of OLS in the multiple regressor case is to minimise $SSE(\beta)$.

The SSE can be written as follows:

$$SSE(\beta) = \sum_{i=1}^{n} y_i^2 - 2\beta' \sum_{i=1}^{n} x_i y_i + \beta' \sum_{i=1}^{n} x_i x_i' \beta \tag{27}$$

If we solve for the first order conditions, we can minimise $SSE(\beta)$. Let $\widehat{\beta}$ be the least squares estimator that minimises $SSE(\beta)$. Taking first order conditions and setting the derivative to zero we obtain:

$$0 = \frac{\partial}{\partial\beta} SSE(\widehat{\beta}) = -2\sum_{i=1}^{n} x_i y_i + 2\sum_{i=1}^{n} x_i x_i' \widehat{\beta} \tag{28}$$

If we then divide by the solution of this by two, we obtain:

$$\sum_{i=1}^{n} x_i x_i' \widehat{\beta} = \sum_{i=1}^{n} x_i y_i \tag{29}$$

---

41 With a single dependent variable, OLS can be calculated by hand and is referred to as Simple Linear Regression. With multiple dependent variables, the OLS becomes a matrix system and is computationally difficult to calculate by hand with larger numbers of dependent variables. Computationally it can be calculated with relative ease in regression software in either Excel, STATA or EViews.

We then can then multiply by the inverse of the term $\sum_{i=1}^{n} x_i x_i'$ in order to isolate $\widehat{\beta}$. The OLS estimator for multiple regressors is given with:

$$\widehat{\beta} = (\sum_{i=1}^{n} x_i x_i')^{-1} (\sum_{i=1}^{n} x_i y_i) \text{ or } \widehat{\beta} = (X'X)^{-1} X'Y \text{ (in matrix form)} \tag{30}$$

This is a system of *k* equations in matrix form. To satisfy this is indeed the minimiser of $SSE(\beta)$, we should satisfy the second order conditions of the system as well. Taking the second derivative we obtain the following second order conditions below:

$$\frac{\partial}{\partial\beta\partial\beta'} SSE(\widehat{\beta}) = 2 \sum_{i=1}^{n} x_i x_i' \widehat{\beta} > 0 \tag{31}$$

As we have a positive definite matrix, the left-hand side of the system will be greater than 0, therefore the second order conditions are satisfied. Therefore, we have seen the derivation OLS estimator that is used in many regression models, including finance regression models that include Fama and French Factors. The next step is to show the assumptions of the OLS estimator and that this estimator is BLUE (Best Linear Unbiased Estimator), by being BLUE, it means this estimator is superior to all other estimators when certain econometric conditions hold in the linear data. If these econometric conditions are violated, the OLS estimator will cease to be BLUE and another estimator may be suitable over the OLS estimator.

OLS is built on a series of 4 assumptions that are as follows:

<u>1. Correct Specification</u>

The actual data generating process must be in a linear functional form.

<u>2. Strict Exogeneity</u>

The error terms of the regression should hold a conditional mean of zero. This in turn means that the error terms should not be correlated with the regressors. If the error terms are correlated with the regressors, they are referred to as endogenous and if they are not they are referred to as exogenous. With endogenous regressors OLS is not valid and Instrumental Variables (IV) may be used instead.

<u>3. Spherical Errors</u>

This assumption is usually split into two parts that of Homoscedasticity and of No Autocorrelation. Homoskedasticity means that the error term has the same variance in each observation, violating this means we have heteroskedasticity instead. OLS would cease to be the best estimator and Weighted Least Squares would be a better estimator. Homoskedasticity can be checked using a scatterplot of residuals versus predicted values. We look for no clear pattern in the distribution of residuals versus predicted values.

There should also be no autocorrelation so that the errors are uncorrelated between observations. This assumption can be violated in terms of time series data or panel data but then again OLS would cease to be the best estimator and Generalised Least Squares (GLS) would be the better estimator.

4. No Linear Dependence

The regressors in X of the system must be linearly independent. This means the matrix X must have a full column rank with almost sure convergence. Finite moments must exist in the regressors up to at least the second moment. If this assumption is violated we have multicollinearity and the value of the regression coefficients of β does not exist.

5. Normality (additional, but not necessary assumption)

It is often assumed that the errors have a normal distribution conditional on the regressors. OLS does not explicitly need this assumption but it does mean that finite sample properties in the area of hypothesis testing can be made. With normally distributed errors, the OLS estimator behaves like the Maximum Likelihood Estimator (MLE). A goodness of fit test such as the Kolmogorov-Smirnov Test can be used to check the normality assumption present in the data.

With these 5 assumptions, OLS is defined correctly and will be BLUE. The definition of BLUE is very much interlinked with the Gauss-Markov Theorem and is not necessary for all 5 assumptions to hold in order to be BLUE. In a linear regression model with errors that have an expectation of zero, errors which are uncorrelated with each other and are homoscedastic means we will be BLUE if the OLS is taken as the estimator. This means that only assumption 3 of Spherical Errors needs to hold for the OLS estimator to be BLUE.

3.2.2 Maximum Likelihood Estimation (MLE)

MLE estimation is employed for the estimation of models. This in turn is true for the class of GARCH models that are used to model volatility in finance. MLE is based on maximising the likelihood function. Suppose that the variable X from $X_1, X_2, X_3, \dots X_n$ has a joint density which is denoted by:

$$f_\theta(x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n | \theta) \tag{32}$$

If we have observed values of $X_1 = x_1, X_2 = x_2, X_3 = x_3, \dots X_n = x_n$, the likelihood of θ is given by the function:

$$lik\ (\theta) = f(x_1, x_2, \dots, x_n | \theta) \tag{33}$$

Therefore $lik\ (\theta)$ is the probability of observing the data we are given as a function of θ. The Maximum Likelihood Estimator (MLE) of θ is the value of θ the maximises $lik\ (\theta)$, i.e. the value of θ that makes the data we are observing the most probable in outcome. If X_i are independent and identically distributed, then the likelihood will simplify to:

$$lik\ (\theta) = \prod_{i=1}^{n} f(x_i | \theta) \tag{34}$$

It can be challenging to maximise a product, we can instead use the fact that the logarithm of this function is an increasing function. Therefore, we can instead maximise the log likelihood and obtain the same result.

$$l(\theta) = \sum_{i=1}^{n} log\, f(x_i | \theta) \tag{35}$$

The advantages of MLE estimation is that MLE estimators can be developed for a large variety of estimators, in particular for models with time varying variance such as the GARCH family. MLE estimators have mathematical and optimality properties that are desirable. As sample sizes increase, MLE estimators become minimum variance unbiased estimators. MLE estimators also have approximate normal distributions and approximate sample variances that can be used for confidence intervals and hypothesis testing.

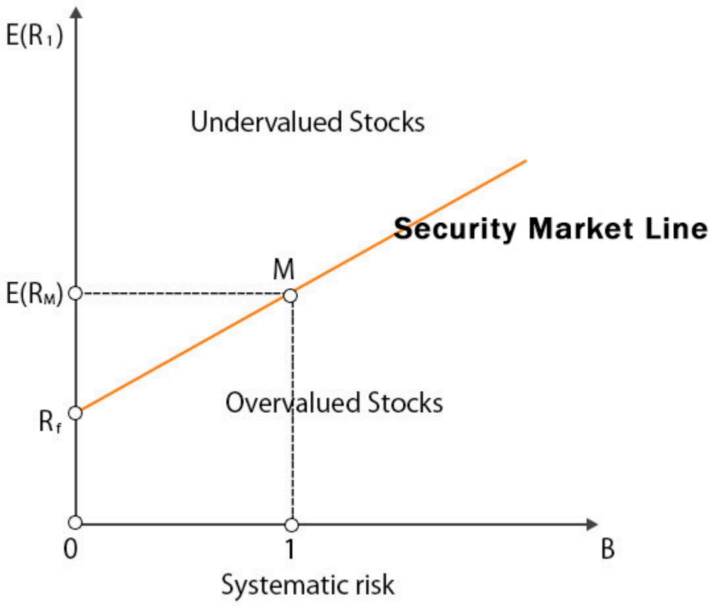
The disadvantages of MLE estimation is that likelihood equations need to be specifically calculated for a given distribution, which makes manual calculation tedious. MLE estimation properties are suited to larger samples and can have bias if samples are too small. MLE estimation can also be sensitive to starting values which are employed in the estimation.

3.3 Theories Employed Regarding Short Interest, Short Selling and Market Returns

The literature review is a very good guide at getting an understanding of past work published in the field and the results obtained from the past work. However, it does a somewhat poor job at explaining relevant explanations to their findings. This section aims to explain the current theories in the field that aim to explain the findings of our literature review regarding short interest, short selling and market returns.

A strong empirical finding is that stocks with high levels of short interest are followed by negative abnormal returns in the long run. This can be seen in Desai et al (2002) and Asquith and Meulbroek (1996) as two relevant examples. This is in contrast with the efficient market hypothesis as we now have public information that is seen to be informative of future stock returns.

Figure 1: Representation of the Security Market Line (SML)



Notes : Data Source from https://www.wallstreetmojo.com/security-market-line. Figure 1 shows the relationship between systematic risk (undiverisifiable risk) and the return of a security.

The efficient market hypothesis states that it is impossible to beat the market because stocks reflect all available information in the stock price, this means that stocks are not going to be under or overvalued and will sit on the security market line (SML). The SML is a representation of the CAPM at work taking systematic risk on the x axis and return on the y axis. An over or undervalued stock would be above or below the SML respectively, while a stock that is fairly priced will sit on the SML itself (see **Figure 1**). The CAPM stipulates that all stocks in theory should sit on the SML and with the backing of the efficient market hypothesis, this theory is said to solidify the CAPM. For a stock to not sit on the SML means the markets are not strong form efficient.

For a stock to be fairly valued, it's price to earnings ratio (P/E ratio) should be equal to its growth rate in earnings per share. This produces a price to earnings to growth ratio (PEG ratio). A fairly valued stock would have a PEG ratio of 1. The P/E is good at gauging how expensive a security is terms of earnings, but it fails to consider the growth rate of earnings, which is important as investors are paying a price today for future earnings. It is very difficult to gauge future expected earnings for a stock; therefore, stock prices move to reflect these changes. Hope, fear and greed from market participants also drives stock prices far away from a fair PEG value. With a financial instrument like a bond, it is easy to judge future cash flows and value the bond correctly, it is much harder to value stocks correctly, especially when markets are not completely rational at times. The role of a good investor is to judge whether a stock is over or undervalued and act upon it by either buying or selling. Thus, Figure 1 above shows a representation of the SML.

Regarding the Efficient Market Hypothesis, three forms of the hypothesis have been proposed. Weak Form Efficiency, Semi Strong Form Efficiency and Strong Form Efficiency. Weak Form Efficiency puts forward that all past information is priced into securities. Fundamental analysis of securities can give an investor with information to produce returns above market averages in the short term but there are no patterns that exist. Thus, fundamental analysis does not provide long-term advantage and technical analysis will not work. Semi Strong Form Efficiency says that neither technical analysis or fundamental analysis can provide an advantage for an investor and that new public information is instantly priced in to securities. Strong Form Efficiency stipulates that all information, both public and private, is priced into stocks and thus no investor can gain advantage over the market as a whole. However, Strong Form Efficiency does not imply that some investors are incapable of capturing high returns but that there are always outliers included in the averages.

The Efficient Market Hypothesis is one of the most debated hypotheses in finance, as many academics and investors argue at the nature of fundamentals influencing stock prices, as opposed to perceptions, the literature findings in regard to short interest and returns shows that the Efficient Market Hypothesis in its strong form at least does not completely hold in this case. The literature in my case has two possible explanations for these negative abnormal returns for high levels of short interest.

The first explanation is referred to as the Overvaluation Hypothesis and the second explanation is referred to as the Information Hypothesis.

3.3.1 Overvaluation Hypothesis

The Overvaluation Hypothesis based on the work of Miller (1977) states that stocks with high levels of short interest are overvalued because investors that are bearish on those stocks are not able to participate in the pricing process. This leaves only optimists participating in the pricing process, who are in a long position in the security. When a stock has a high level of short interest, borrowing the stock to short can be expensive, this leads to stock overpricing and thus lower returns.

The Overvaluation Hypothesis shows a market efficiency problem, where a set of market participants are excluded due to the expensive nature of borrowing the stock. In efficient markets there should be no problem finding a buyer or seller, but a breakdown of market efficiency can lead to this problem.

This overvaluation is evident during short-sale bans, where short sellers are restricted from participating in the pricing process. This can lead to dramatic price changes once the short selling ban is over as many short sellers try to take advantage of this overvaluation. Rather than a gradual floor being found in a stock price, this can lead to dramatic volatility as a vast degree of short sellers suddenly enter the market. It may be worth considering a more gradual re-entry of short sellers into the pricing process if a short sell ban is again implemented on certain markets.

3.3.2 Information Hypothesis

The Information Hypothesis builds on the notion that short sellers are informed traders, which has been widely documented in the literature. Short sellers are seen by Gutfleish and Atzil (2004) to be traders with very good analytical skills and take on short positions when public information allows them to see low fundamental valuations. Short sellers have been known to examine cash flow statements and watch for elevated price earnings ratios when certain sectors are in decline in order to take advantage. As a result of this information processing short sellers are able to identify stocks which are likely to show weakness in the future at a very early stage. This allows us to see the relationship with stocks with high short interest and future negative abnormal returns.

Akbas et al. (2008) discriminate between the two theories of overvaluation and information and finds support for the latter in their study, though it is likely that both hypotheses have elements of play in determining exactly why high short interest stocks have lower future returns. It is worth noting that future returns extend into the long term, while in the short term high short interest stocks are at very high risk of a short squeeze just on overall macro market sentiment rather than firm specific sentiment.

The debate between the Overvaluation Hypothesis and Information Hypothesis remains, however more studies such as Akbas et al. (2008) have found overwhelming support for the Information Hypothesis than the Overvaluation Hypothesis and given the current state of play in the literature, this seems to be the case going into the near future.

3.4 Theories Employed Regarding Short Selling and the Relationship with Liquidity, Price Discovery and Volatility

This section looks at the theories behind the findings of short selling bans and the relationship with liquidity, price discovery and volatility. The best way to explore the effect of short selling on these market indicators is often to remove short selling from the market, naturally no academic study could force a market to do that in the name of an investigation but there have been many cases in the past where short selling has been banned due to exceptional market conditions.

Short selling bans have most often occurred during periods of extreme financial turmoil such as the 2007-2009 Financial Crisis and the 2015 Chinese Currency Devaluation. These periods allow me to explore the effect of short selling on these market indicators. A regularity authority is responsible for the short sale ban, this can take the form of the FCA (financial conduct authority) in the UK and the SEC (securities and exchange commission) in the US. Each country will have their own regulatory authority responsible for the regulation and stability of their financial market. Short selling bans are usually of a temporary nature and can apply to all stocks or a certain sector of stocks. Short selling bans can also target naked short selling or all short sales, naked short selling being the selling of a security without first borrowing the stock from the broker. It is up to regulatory authorities to decide what type of short selling to target.

3.4.1 Volume Hypothesis

One of the most common understandings is that a short selling ban removes a great deal of potential volume in the market. Some market participants are exclusively short sellers, while others are not prepared to trade long at all in a falling market.[42] A decrease in volume will widen bid-ask spreads and make the buying and selling of securities much harder, hence the decrease in liquidity.

Volume and volatility are one of the most studied relationships in finance and it has been found that a decrease in volume automatically causes an increase in volatility in some studies. The two prominent hypotheses behind the volume-volatility relationship are the Mixture of Distributions Hypothesis introduced by Clark (1973) and the Sequential Arrival of Information Hypothesis introduced by Copeland (1976).

The Mixture of Distributions Hypothesis assumes that the volume-volatility relationship is dependent on the information flow rate in the market. The hypothesis states that the positive relation between volume and volatility is due to all traders simultaneously receiving new information. The Sequential Arrival of Information Hypothesis assumes that traders receive new information in a sequential form, so that traders change their position as new information enters the market.

Therefore, the decrease in short selling volume leads to an increase in volatility, which this theory is also backed up by the volume-volatility literature as well, though some studies argue against this showing volume has little effect on volatility. The price discovery mechanism also links into the decrease in liquidity as one potential factor, as liquidity decreases, price discovery becomes harder. Therefore, less short selling means less liquidity and thus a decrease in the price discovery mechanism.

Financial regulatory authorities such as the SEC in the US and FCA in the UK can have influence on the volume that enters a market from outright bans (such as short selling bans) or suspending securities entirely from trading (halting a security). Trading hours can also affect volumes with the highest volumes usually recorded in the first 30 minutes of the trading day in financial markets. Volumes also rise during crossover periods where multiple markets are open, such as the 2:30-4pm GMT crossover between the US and European Markets.

Many stock exchanges such as the NYSE offer trading that is extended hours outside of the normal trading day, what studies have observed is far less volume in extended hours trading leading to more volatility and wider bid-ask spreads. The NYSE states that extended hours let investors trade on news that is often announced after the close of the NYSE trading day, such as earnings reports and federal reserve decisions. However, critics argue that extended hours are not necessary and add unnecessary fear into a market open if there is some sort of crisis occurring in world financial markets. Given investors have nearly 7 hours to

[42] It is worth mentioning that a falling market is also a very fast-moving market and prices tend to gap rather than incrementally decrease. This can lead to substantial gains for short sellers and can be very dangerous to form a long position until a bottom is found and the buyers and sellers have stabilized. Panic selling and stop losses on margin, contribute to the fast nature of falling markets. Markets are seen to be asymmetric, with falls faster than rises, often colloquially referred to stock movements as taking the stairs up and the elevator down.

buy or sell securities over 5/7 days, the advent of extended hours may not be necessary unless investors are also a market maker where they can benefit from the large bid-ask spreads.

3.5 Theories Employed Regarding Short Selling and Market Manipulation

The literature of whether short sellers are manipulating the market or not rests on whether short sellers are trading on public information or whether they are trading on noise. If short sellers are trading on public information, they are said to be informed, while if they are trading on noise, they are said to be uninformed. Market manipulation is a serious offense and the SEC and FCA prosecute individuals engaging in it. Disclosure of information is very important for a crime to be considered market manipulation, trading on non-public information can constitute market manipulation. There have also been governmental debates on the nature of public information and whether non-official sources such as Twitter or Facebook constitute as public information or not. Officially the SEC recognises Twitter as a source of public information if investors are fully aware that it is a source of information for a particular security, this includes Twitter accounts held by the company and its board of directors.

3.5.1 Informed Traders Hypothesis

Boehmer et al. (2008) and Engelberg et al. (2012) are both papers which have reported that short sellers are indeed informed and are trading on information rather than noise. If short sellers are informed, they cannot be market manipulators since market manipulators would not be trading on good public information. Therefore, this hypothesis holds on the premise that short sellers have no real contribution as a whole to market manipulation since they are informed. Removing short sellers will not cause a reduction in market manipulation, since the majority of them are not responsible for it.

The ability of short sellers to predict the future direction of stocks shows whether they are informed or not. High short interest ratio stocks underperforming compared to low short interest ratio stocks is another good indicator that the majority of short sellers may be informed. Therefore, the short sellers are able to price in future events that may not be predictable with absolute certainty.

It is also worth noting that large capitalisation stocks experience large degrees of volume and are thus much harder to manipulate in price as opposed to very small capitalisation stocks. The fact that the majority of short sales occur on large capitalisation stocks over very small capitalisation stocks, also suggests that the majority of short sellers are informed and trading on good public information. Short sellers target a broad range of stocks from large capitalisation, mid capitalisation to small capitalisation.

3.6 Conclusion

In conclusion, we have seen both the models and methods employed in this thesis and the theories behind the findings of our literature review. We started off with the Capital Asset Pricing Model, which has been the building block of most asset pricing models. We have seen how the CAPM puts emphasis on systematic risk as whole and sees systematic risk as the prime contributor to return. Over the years we have learnt that this is not truly the case and other components are clearly contributing to the return of a security or portfolio.

We have seen the development of the Fama and French Three Factor Model, how Fama and French (1993) formed their model from portfolios built on size and book to market ratio. Each of the components of the Fama and French Three Factor Model is constructed from these portfolios or the market risk premium. We were able to observe the robustness of the model in testing and see how it is a much better fit than the CAPM. We have seen how Carhart (1997) extended the Fama and French Three Factor Model to include momentum, which up until recently has been the standard in asset pricing models.

Subsequently premiums of investment and profitability began to show up in studies and this pushed Fama and French to develop their Fama and French Five Factor Model. Although the debate remains on the exclusion of the momentum factor, which has been seen as a key component of asset pricing models. We see that the construction of the Fama and French Five Factor Model is much more complex, with 18 portfolios in total being used. The Fama and French Five Factor Model has been shown to outperform the Fama and French Three Factor Model in various studies.

We then see the construction of the GARCH (p,q) model and how it uses an ARMA model as the variance of the error term. We see how GARCH models are particularly useful for volatility clustering. We then see how bid-ask spreads are created and how a bid-ask model can be used to model liquidity in financial markets. We see how bid-ask spreads can be difficult to obtain and a good degree of estimators exist, we in particular are interested in the Corwin and Schultz (2012) estimator.

Moving on we look at the methods employed of OLS and MLE, looking at the derivation of both estimators and their relevance to the research questions that are being used. We see a set of conditions that need to hold for OLS to be BLUE.

Next, we see the theories behind the literature review, we see two prominent theories regarding short interest and abnormal negative returns. The first being the Overvaluation Hypothesis and the second being the Information Hypothesis. The Information Hypothesis also helps to explain how short sellers are indeed informed and therefore do not contribute to market manipulation. We also see how decreases in volume, affects liquidity, volatility and price discovery. We see that ample volume is necessary for the smooth functioning of a market and a lack of volume can have negative effects on liquidity, volatility and price discovery.

Overall this chapter outlines the theory behind our methodology. I will be employing the models and methods explained in this chapter in my research and now we have a good understanding of how these model and methods have been constructed and their limitations in this study. The theories regarding short interest we have seen here allow to see the possible reasons to the findings of our literature review and whether we are likely to encounter similar circumstances in my study.

CHAPTER 4: DATA DESCRIPTION

The data description chapter lays out the data I am using for my research questions. This is crucial as it allows the reader to see my data sources and to find any potential problems in the study from the source rather than working back from the results. It also means that my study can be replicated and this is essential for any empirical research project. The ability to replicate a project will increase the reliability of the results, this is notably different from the accuracy of a study, which can only be increased by each individual research project being more cautious in its collection of data, production of results and conclusions.

This research project is focusing on four questions so we have done an individual data description for each question, as the datasets are unique to that particular question. The first three research questions are based on a similar methodology, but have unique datasets. The last one research question uses unique methodologies and again uses a unique dataset.

4.1 Five Factor Model as an Adjuster for Risk Premium (Research Question 1)

My original research question states the following:

In adjusting for risk, Boehmer et al. (2010) uses the Fama and French (1993) three-factor model augmented by the momentum factor. However, is this long/short strategy still valid if a different and more recent model, such as Fama and French (2015) five-factor model, is used to adjust for risk premium?

For this research question, I will be following the standards of Boehmer et al. (2010) as it will substantially help in the comparison of both models. I will be examining monthly short interest data from January 2001 to January 2010 in stocks listed on the NYSE and Nasdaq exchanges. These are the 2 largest exchanges by market capitalisation in the United States. I am using this time period as it was part of the one explored by Boehmer et al. (2010) and I want to make an as direct comparison as possible to represent the US market. The datasets constructed by Boehmer et al. (2010) are not freely available to the public. I will be nevertheless be creating my own datasets using Kenneth French's Data Library and Thomson Reuters DataStream. It is understandable why these datasets are not open to the public, as they would have taken considerable time and skill to create, though my investigation would have been hugely aided as a means of reference if they were. Even with pre-formed datasets, it would have been better to construct my own for replicability and robustness. Bradford D. Jordan of the University of Kentucky [43], who is the corresponding author of Boehmer et al. (2010) was contacted in regard of this study, but unfortunately was not able to provide datasets as a reference. However, the publication of Boehmer et al. (2010) is a quality in depth study with the fact that the findings have not been invalidated for a period of 10 years, means I can have confidence in the robustness of their study. In other words, it is not solid to just look at years of revisions as reviewers can

[43] Full details: Gatton College of Business and Economics, University of Kentucky, Lexington, USA, email: bjordan@uky.edu

make the same mistake evaluating a dataset with flaws or a methodology with flaws. Therefore, it is better to see a longer time period of invalidation in the robustness of a study. Their findings have also been extremely similar to Desai et al. (2002) and Asquith et al. (2005), this further gives me great confidence in their results. Further on the results of Boehmer et al. (2010) are consistent with the previous literature in regards to the efficient nature of short sellers in the market and the overperformance of lightly shorted stocks over heavily shorted stocks.

My observational dataset will consist of monthly stock returns from January 2001 to January 2010 in line with Boehmer et al. (2010). I will be including stocks that have been in the market from January 2001 to January 2010. My monthly stock return provider will be Thomson Reuters DataStream. I also follow Boehmer et al. (2010) and impose the sanction that the stock is listed for at least 1 year and this removes all IPOs (initial public offering). I don't impose any other restrictions on my stock dataset.

I will download the relevant factor data for the Five Factor Model from the data library of Kenneth French. This gives the factor data for the Five Factor Model in weekly and monthly form. In my case we are interested in the monthly data. Kenneth French keeps his data updated regularly so it is possible to obtain data for his factor portfolios up to 2 months before the current date usually.

The data from Kenneth French's library will include the factor loadings for each month attributed to each of the factors in the Fama and French (2015) Five Factor Model that is often used as a benchmark for asset pricing securities. Kenneth French's library also keeps a weekly and monthly listing of the US 3-month treasury bill return, which is often taken as the risk-free rate [44]. Therefore, my data regarding the risk-free rate is also available from his library. The factor data I will be using will include the excess return on the market factor, the small firm effect factor (SMB), the book value factor (HML), the profitability factor (RMW) and the investment factor (CMA).

The factor loadings are formed from portfolios. The SMB factor is formed by the average return on nine small stock portfolios minus the average return on nine large stock portfolios. The HML factor is formed by the average return on two value portfolios minus the average return on two growth portfolios. The RMW factor is formed by the average return on two robust profitability portfolios minus the average return on two weak profitability portfolios. The CMA factor is formed on the average return on two conservative investment portfolios minus the average return of two aggressive investment portfolios. The excess return of market factor is calculated by using the return of the market (taken as all firms listed on the NYSE (New York Stock Exchange), Amex (American Stock Exchange) or Nasdaq with CRSP share code 10 or 11) minus the US 3-month treasury bill rate.

My Short Interest Data will be for January 2001 to January 2010 time period and will be in monthly form. This data will come from Thomson Reuters DataStream and will indicate the level of short interest in each particular stock. My aim is to follow Boehmer et al. (2010) and form portfolios of stocks each month based

44 The risk-free rate is the rate of return which is effectively considered with near absolutely no risk, with the emergence of the US as a dominant superpower it is often taken as the 3-month US Treasury Bill. However, this is not set in concrete and other authors may use a different standard for a risk-free rate. In my study I am using the 3-month US Treasury Bill. My risk-free rate is correlated with the federal reserve discount window rate (the rate at which the United States Federal Reserve lends to authorised banks).

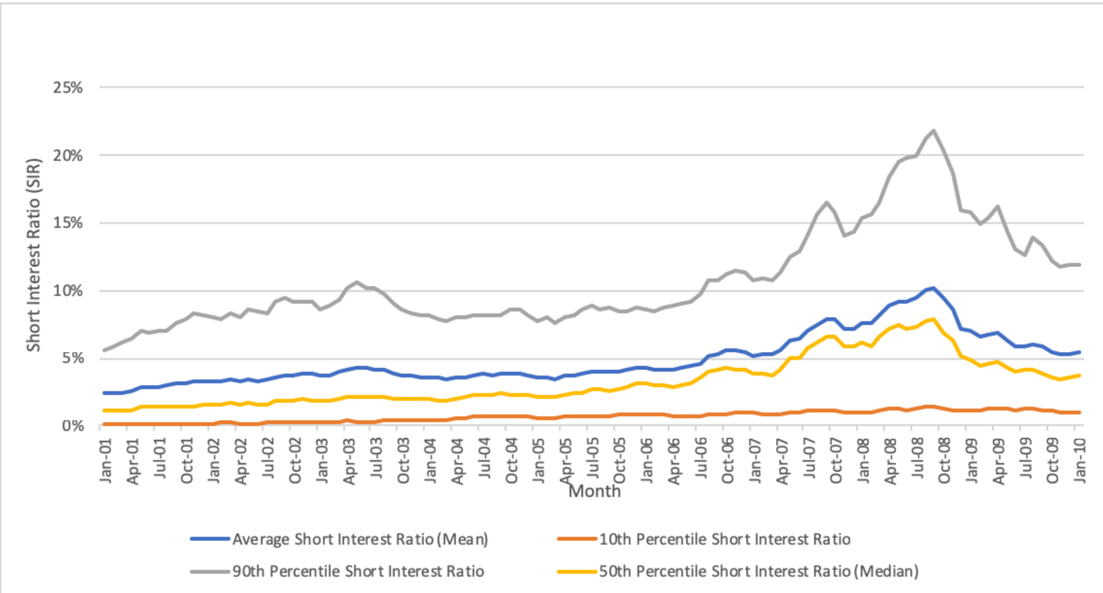
on short interest levels and run a time series regression on the monthly portfolio excess return using the Five Factor Model. More details regarding this can be found in the methodology chapter.

4.1.1 Dataset Statistics for the United States

My dataset consists of 109 monthly observations for the Thomson Reuters US Total Return Index between the months of January 2001 and January 2010 inclusive. The Thomson Reuters US Total Return Index attempts to model to return of the US economy and consists of stocks listed on the NYSE and Nasdaq. On average I have around 1700 firms per month in each of the monthly observations.

I am particularly interested in the short interest that my dataset encompasses as a whole, the level at which the short interest changes each month for the entire dataset and the dispersion of short interest between securities in the dataset.

Figure 2: Distribution of the Short Interest Ratio in the Thomson Reuters US Total Return Index



Notes: Data Source using Microsoft Excel from Thomson Reuters US Total Return Index DataStream Data. Figure 2 shows the distribution of the short interest ratio in the Thomson Reuters US Total Return Index from January 2001 to January 2010. The short interest ratio is calculated as the number of shares sold short divided by the total number of shares in the open market. The Thomson Reuters US Total Return Index encompasses our whole dataset.

Figure 2 shows the distribution of the short interest ratio in the Thomson Reuters US Total Return Index. We see that the short interest ratio is gradually increasing across our dataset between 2001 and 2007, until a very large spike in 2008 attributed to the global financial crisis. The average short interest ratio (mean) is shown with the blue line in figure 2 and is calculated by taking the average of all the short interest ratios of each security for a particular month. Each month securities enter and leave the Thomson Reuters US Total Return Index and thus the short interest ratio for the index will change each month.

The average short interest ratio moves from 2.5% in early 2001 to a peak of 10% in mid 2008 and comes back down to 5% in early 2010, this is in line with short interest levels reported by Desai et al. (2002) and Boehmer et al. (2010). Boehmer et al. (2010) reported a dataset average short interest ratio level of around 4% in 2005, in line with our findings.

The other three lines of significance are the 50th, 10th and 90th percentile lines of the short interest ratio. The 50th percentile line tells us the short interest ratio at which 50% of all stocks in each monthly sample are below. The 10th percentile line tells us the short interest ratio at which 10% of all stocks in each monthly sample are below. The 90th percentile line tells us the short interest ratio at which 90% of all stocks in each monthly sample are below. The 50th percentile line is more commonly referred to as the median. We can immediately see that the data generally each month in this dataset is skewed towards the lower short interest ratios, this is due to the fact that the mean short interest ratio is above the median short interest ratio. This is a natural finding since the majority of stocks will have good or great financial health while the minority will have average or poor financial health.

The pick-up in short interest is evident during the global financial crisis, with the average short interest ratio hitting 10% and the 90th percentile short interest ratio hitting 22%. These are clearly exceptional circumstances, with a high level of short interest. Short selling was banned in certain financial stocks during the global financial crisis, but the majority of stocks were allowed to be sold short. This certainly put extra downward pressure on stocks and was evident since the Dow Jones Industrial Average moved from 14,198 in October 2007 to 6443 in March 2009. This is close to a 54% decline from peak to trough (excluding dividends).

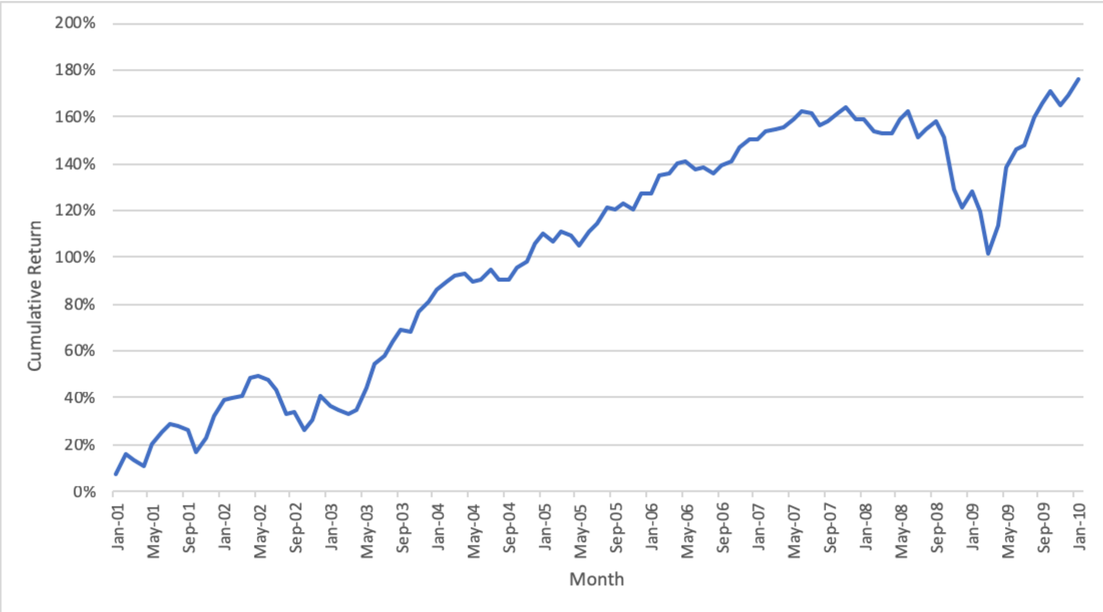
What also is evident is the rally that was to follow the drop, the rally was very quick in nature and this is to be attributed to a combination of short squeezes and the large degree of money on the side-lines entering the market. A lot of short positions would have been forced to buy, once the market turned and this amplified the buying. The high levels of short interest indicate a large degree of fast short position covering would have taken place once the market turned. It was a case of short sellers shorting every bear market rally, until a series of rallies took place and short sellers realised that a technical bull market was near formation.

From this sharp fall and rise we can attain that fear and greed [45] are one of the key drivers of short-term market movements. These are often measured using market sentiment indicators of which one is short interest. Other indicators include the put/call ratio, market momentum, market volatility, safe haven demand, junk bond demand and stock price breadth. A high short interest is often an indicator of negative market sentiment, which was the consensus in the global financial crisis. If stocks are overvalued relative to the earnings they may generate in the future, more and more short sellers would be willing to open a short position knowing this. Thus, driving up levels of short interest.

45 Fear and greed are often one of the key drivers of short-term market movements. If investors insist on trying to time their participation in equities, they should try to be fearful when others are greedy and greedy only when others are fearful (Buffett, 2004). The point of maximum greed is the point of maximum risk and the point of maximum pessimism is the point of maximum opportunity. Having a contrarian mentality always helps in investing rather than following the herd. Widespread fear in the market is seen as an investor's friend, since it serves up bargain prices relative to future cash flows in a business, personal fear is seen as an investor's enemy, as it leads to irrational decisions such as selling into a market crash.

Looking at the extremes, the 90th percentile of short interest ratios will be below 10% on average. The 10th percentile of short interest ratios will be below 0.5% on average. If an investor does find a company in the extreme percentiles, it is best to check the balance sheet and income statements as this is often indicative of something very well or very wrong with the financial health of the company. Some key indicators of problems in a company can be a low debt to equity ratio (below 1) in the balance sheet, suspension or cuts in dividend payments and falling year on year revenues in the income statement. Conversely in good companies the opposite may hold true, where we have a high debt to equity ratio (above 1.5), raises in dividend payments with adequate dividend cover (above 1) and rising year on year revenues.

Figure 3: Cumulative Return of the Thomson Reuters US Total Return Index



Notes: Data Source using Microsoft Excel from Thomson Reuters US Total Return Index DataStream Data. Figure 3 shows the cumulative return of the Thomson Reuters US Total Return Index, which is my dataset. The return is familiar with the S&P 500 over the period of 2001 to 2010 on a total return basis, which is to be expected as the S&P 500 is a sub component of my dataset. I post returns on a monthly basis and this affects the volatility in the chart. Daily returns would show more volatility as I have more data points.

Looking at **Figure 3** is clear that on a total return basis (this means we include dividends reinvested on top of capital gains as return) it has been wise to be invested in the American economy between 2001 and 2010. The only way an investor would have lost money is if he or she invested at the peak of bull market in 2007,[46] only to be driven by fear and pull his or her investment out of the market in 2009. This is another case of why it is evident to have a long-term horizon in investing and for an investor not to invest in the market if their investment period is not greater than 5 years, the fluctuations of the market can be large over a time period less than 1 year, especially in cases of extreme fear in the market. We can see that the total return is negative for the 2007 to 2009 period as many corporations encountered restructuring and either cut or

46 Investors can get carried away and buy equities when valuations don't reflect future earnings growth. This is the point of maximum risk and should be avoided. As famed investor Benjamin Graham would say, "buy stocks like you buy your groceries, not how you buy your perfume" (Graham, 1949). When investors fail to ask "how much" is when they run into trouble. The tech bubble is a good example where investors never asked the price of a hot stock, the consequences were devastating, as price/earnings multiples contracted to reflect the lack of future earnings growth. Valuation matters eventually and the market corrects overpricing.

suspended their dividend payments.[47] Many academic papers such as Hauser (2013) and Floyd et al. (2015) document these dividend cuts and restructurings as a result of effects of the 2007-2009 financial crisis.

Another point worth mentioning is that the bear market of 2002 to 2003 was not severe in nature, with the total return remaining relatively flat to slightly negative. This adds to the negative economic significance of the global financial crisis, as a normal bear market would not have been as severe. The only comparable crash to the global financial crisis would be the 1929 Wall Street Crash. These are often once or twice in a lifetime event on the macro economy. My dataset includes three bull markets and two bear markets, so will be indicative of the natural business cycle.

4.2 Is There Arbitrage? (Research Question 2)

Our original research question states the following:

After the publication of Boehmer et al. (2010), does the opportunity for excess returns remain, or have investors adopted this strategy and arbitraged away the excess returns?

The dataset for this research question will involve stocks and stock returns in the United States from 2010 to 2017, for the period after the publication of Boehmer et al. (2010). I am again particularly interested in monthly stock returns that represent the US market for this period and so I will be using the Thomson Reuters US Total Return Index as my source for these returns. The Thomson Reuters US Total Return Index includes stocks listed on the NYSE and Nasdaq that are most representative of the US economy as a whole. The total return nature of the index means that dividends are also included in the monthly returns, which is crucial as dividends at times have made up of 50% of the return that investors as a whole has received from investing in the markets.

If investors were to adopt a strategy as such that is proposed by Boehmer et al. (2010), they would be interested in the total return they are receiving rather than just the capital gain of the securities in question. The Thomson Reuters US Total Return Index will include stocks that in general are representative of the US economy as a whole.

I also impose the same sanctions as my first research question in that the stocks must be listed for at least 1 year, ruling out any IPOs [48] and clearly be listed on either the NYSE (New York Stock Exchange) and Nasdaq. I don't apply any other restrictions.

The original model used in Boehmer et al. (2010) was the Fama and French Three Factor Model augmented with the Momentum Factor, also known in the literature as the Carhart Four Factor Model. Therefore, in this case I will be using the factor data in the Kenneth French library in monthly form for the Fama and

---

[47] When companies are in financial distress, they must suspend dividend payments to shareholders and focus on paying creditors. Failing to pay creditors can lead to the fire sale of assets in administration or outright bankruptcy. It is often advised to pay down excess debt over recommencing dividend payment.

[48] IPOs or Initial Public Offerings can be very tricky in regard to data. With IPOs there is a sustained level of high volume initially and restrictions on selling by insiders which can affect short interest levels. It is always best to let a security stabilise in the market before any research is conducted on it.

French Three Factor Model and also obtaining the Momentum data from his same library. This data is formed in the same way as the Five Factor Model Data by taking portfolios and averaging returns.

The Fama and French Three Factor Model is a simpler model than the Fama and French Five Factor Model, as it involves less components of which have had more rigorous testing over the years. The core of both models is based on the excess return of the market multiplied by the beta of the security (i.e. the CAPM) but both extend the CAPM further to explain returns that the CAPM often failed to explain over time.

The factor data required from the Kenneth French library will include the factors for the excess return of the market, HML and SMB. The factors of the excess return of the market, high price to book vs low book price to book and the small firm effect from the three factors of the Fama French Three Factor Model. The momentum factor is also found in the library and completes the model to form the Carhart Four Factor Model.

For this research question I could have easily implemented the updated Five Factor Model, but this would have made my comparison less stringent, given that Boehmer et al. (2010) employed the Carhart Four Factor Model. Though it would be worthwhile as a future project to test out this dataset using an updated model as to reflect the fair rate of asset return given the risk taken as possible.

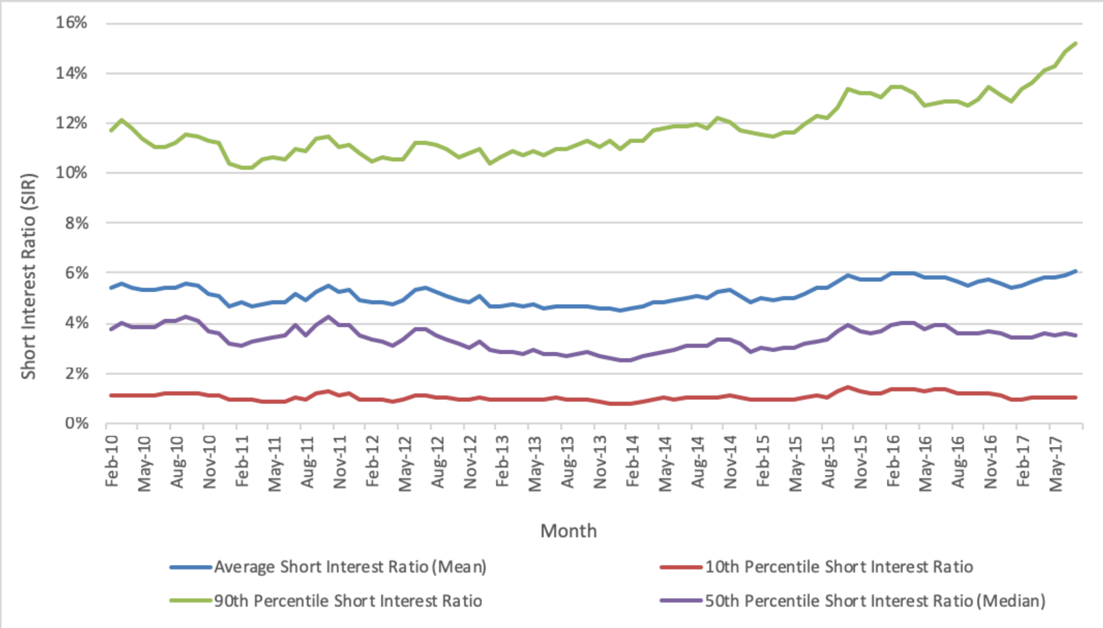
I will again be using monthly short interest data published at the end of the month for stocks that is published in Thomson Reuters DataStream for the period of February 2010 to July 2017, this monthly short interest data will allow me to form our portfolios at the start each month and rebalance them as needed.

4.2.1 Dataset Statistics for the United States

My dataset consists of 90 monthly observations for the Thomson Reuters US Total Return Index between the months of February 2010 and July 2017 inclusive. The Thomson Reuters US Total Return Index attempts to model to return of the US economy and consists of stocks listed on the NYSE and Nasdaq. On average I have around 2000 firms per month in each of the monthly observations.

I am particularly interested in the short interest that my dataset encompasses as a whole, the level at which the short interest changes each month for the entire dataset and the dispersion of short interest between securities in the dataset.

Figure 4: Distribution of the Short Interest Ratio in the Thomson Reuters US Total Return Index



Notes: Data Source using Microsoft Excel from Thomson Reuters US Total Return Index DataStream Data. Figure 4 shows the distribution of the short interest ratio in the Thomson Reuters US Total Return Index from February 2010 to July 2017. The short interest ratio is calculated as the number of shares sold short divided by the total number of shares in the open market. The Thomson Reuters US Total Return Index encompasses my whole dataset.

Figure 4 shows the distribution of the short interest ratio in the Thomson Reuters US Total Return Index. I see that the short interest ratio is relatively stable across our dataset, the average short interest ratio (mean) is shown with the blue line in figure 4 and is calculated by taking the average of all the short interest ratios of each security for a particular month. Each month securities enter and leave the Thomson Reuters US Total Return Index and thus the short interest ratio for the index will change each month.

The average short interest ratio is stable around the 5% mark, this is in line with short interest levels reported by Desai et al. (2002) and Boehmer et al. (2010). Boehmer et al. (2010) reported a dataset average short interest ratio level of around 4% in 2005, the highest short interest ratio in his entire dataset. Boehmer et al. (2010) gradually saw the average short interest ratio rise from 1% to 4% from the year 1988 to 2005, showing the growing increase in shorting activity in the markets. My average short interest ratio is higher than Boehmer et al. (2010) and this is to be expected.

We do see a minor dip in the average short interest ratio in the year 2013, which is of particular interest as 2013 was a very strong year in returns for the markets. This shows a good example of a correlation between the returns of the markets and shorting activity, when market returns are good, shorting activity is subdued to an extent. The average short interest ratio after 2013 picks up again and attains new highs in 2017, I expect shorting activity to increase towards the end of this 2009 bull market and into the start of a new bear market.[49]

49 A new bear market is classified as a drop of 20% or more from the all-time highs of the previous bull market, a drop of less than 20% is classified as a correction and any regain later constitutes the same bull market. The name "bear market" and "bull market" come from the behaviour of those animals. A bear is said to attack with its paws in a downward fashion, while a bull attacks with its horns in an upward fashion. The downward and upward fashions represent the general market trend.

The other three lines of significance are the 50th, 10th and 90th percentile lines of the short interest ratio. The 50th percentile line tells us the short interest ratio at which 50% of all stocks in each monthly sample are below. The 10th percentile line tells us the short interest ratio at which 10% of all stocks in each monthly sample are below. The 90th percentile line tells us the short interest ratio at which 90% of all stocks in each monthly sample are below. The 50th percentile line is more commonly referred to as the median.[50] We can immediately see that the data generally each month is skewed towards the lower short interest ratios, this is due to the fact that the mean short interest ratio is above the median short interest ratio. This is a natural finding since the majority of stocks will have good or great financial health while the minority will have average or poor financial health.

If the mean and the median short interest ratios were equal it would indicate that the data has no skew and would be like a normal distribution. The mean above the median indicates the data is skewed towards the lower figures, while the mean below the median indicates the data is skewed towards the higher figures.

Looking at the extremes of the 10th and 90th percentile lines we can see that the short interest ratio will rarely be found above the 15% mark and found rarely below the 1% mark, when it is found to be above and below these marks it is worth noting the nature of the company, as truly extreme short interest ratios indicate something unique in the fundamentals of the company. Usually a company exhibiting very high short interest is in more of a poor condition balance sheet wise and income statement wise, while a company with a very low short interest ratio is usually of very good health balance sheet and income statement wise. This has been taken by a case by case basis but this is what I have observed conducting this research on this data so far.

This relationship again shows that short sellers may be efficient in targeting the weakest of stocks, while foregoing the strongest of stocks. Short sellers may be opening short positions while looking at fundamentals of companies, rather than randomly selecting stocks to short.

What is also of interest is that the behaviour of the 90th percentile is very different to the others, in particular after the 2015 time period. We see a rise in the 90th percentile short interest in excess of the other lines in Figure 4. This is an indication of stocks in the 90th percentile having more extreme levels of short interest compared to before, this could be due to an increase in short sellers in the market as seen in Boehmer et al. (2010) and Desai et al. (2002), which are targeting the stocks of lower quality (poor financial health). Extreme high short interest stocks often include stocks categorised as penny stocks, in particular as seen in Liu et al (2015) where the average short interest ratio of penny stocks is 14%. Short interest levels above 10% are considered high.

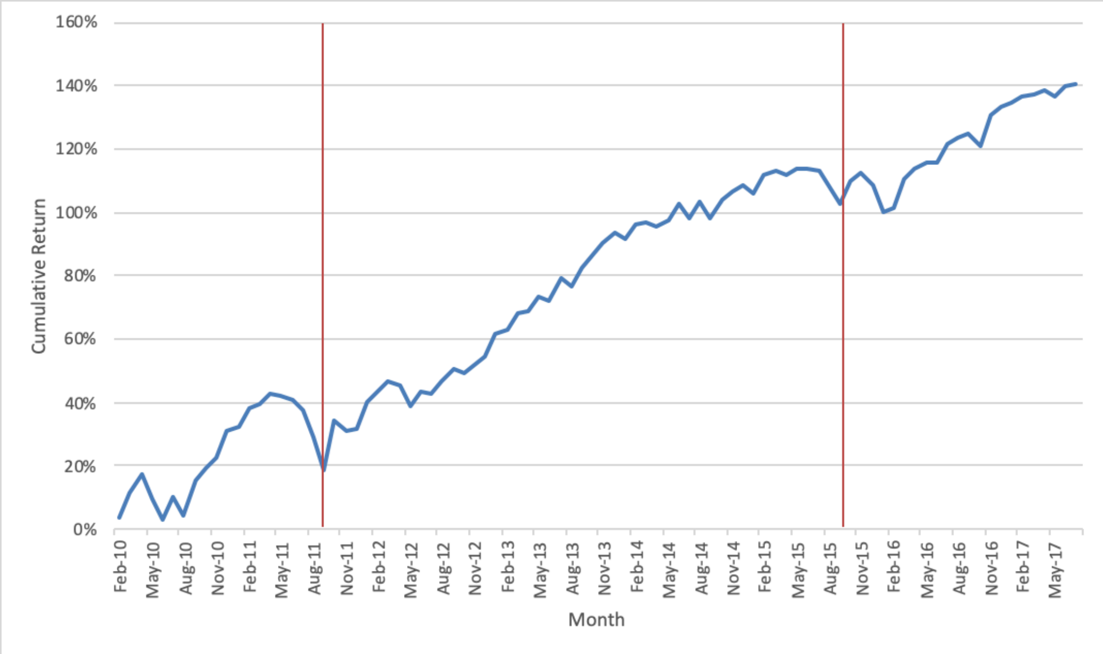
It could also be an indication of deterioration of quality in stocks issued, in particular in the IPO market as seen in Schultz (2003) where IPOs cluster around the top of bull markets. It is often easiest to conduct poor quality IPOs at higher valuations when market conditions are favourable such as in bull markets (where there is ample liquidity) as pointed out in Ritter and Welch (2002). This could lead to an increase in short interest

---

50 It is often debated whether to report the mean or median in statistical analysis, for data with high skew, the median is often a better judge of the central tendency of the distribution. I however report both as needed.

on these stocks (as short sellers in general by the literature are seen to be efficient and will target overvalued stocks).

Figure 5: Cumulative Return of the Thomson Reuters US Total Return Index



Notes: Data Source using Microsoft Excel from Thomson Reuters US Total Return Index DataStream Data. Figure 5 shows the cumulative return of the Thomson Reuters US Total Return Index, which is my dataset. Two vertical red lines indicate the 2011 Eurozone and 2015 China market crisis' respectively which lead to a fall in the total return of the index.

Figure 5 shows the cumulative return of the Thomson Reuters US Total Return Index. Looking at Figure 5 we can see that the return is noticeably similar with that of the S&P 500 over the last seven years, which is not surprising since the S&P 500 [51] is a sub component of our dataset. However, given that we have used monthly returns for the dataset rather than daily returns, we see less volatility in the chart as a whole. The smaller time periods would allow for the complete volatility of the dataset to come through.

The two major dips in the dataset (marked with red vertical lines respectively) are in 2011 and 2015, 2011 being fears of the European Sovereign Debt Crisis and 2015 being fears of Chinese growth deceleration and Yuan devaluation. These are global events and would affect the whole world economy, thus it is natural for a US centric index to react negatively to them. Between February 2010 and July 2017, we have been in a strong bull market in the US and this dataset on a cumulative return basis reflects that.

My dataset is one of few that is measured across a complete bull market, Desai et al. (2002) measured from 1998 to 2004 across the end of the dot com bubble and into the following bear market. Boehmer et al. (2010) measured across 1988 to 2005 across the 90s-bull market, the dot com bubble, the following bear market and the subsequent bull market. The behaviour of volatility and returns are different across bull and bear

51 The S&P 500 is a value weighted index of the largest 500 companies trading on US exchanges, it is said to be a snapshot representation of the US economy as a whole, however better representations exist that incorporate more firms such as the Thomson Reuters US Total Return Index. The S&P 500 is composed mostly of large capitalisation firms, so a lot of mid capitalisation and small capitalisation exposure is lost.

markets and this needs to be noted. Volatility generally rises in a bear market, while returns are negative across the period.

Table 7: Distribution of Return in the Thomson Reuters US Total Return Index by Year

	(1)	(2)	(3)
Year	Average Monthly Return	Average Monthly Standard Deviation	Average Monthly Risk-Free Rate
2017	1.1%	1.5%	0.1%
2016	2%	5.1%	0%
2015	0%	4%	0%
2014	1.3%	3.6%	0%
2013	3.2%	3.1%	0%
2012	1.9%	3.8%	0%
2011	0.1%	6.6%	0%
2010	2.8%	6.7%	0%

Notes: Data Source using Microsoft Excel from Thomson Reuters US Total Return Index DataStream Data. Table 7 shows more specifically where the returns are located in the dataset. Data author's own calculation.

Table 7 shows the distribution of return in the Thomson Reuters US Total Return Index by year. Looking at Column (1) in Table 7 we can see that the monthly returns vary between 0% and 3.2% on average, with 2015 being the worst year for the dataset and 2013 being the best year for the dataset. The standard deviation in Column (2), which is a measure of the variation of returns in my case, shows that 2010 and 2011 were some of the most volatile months, while 2017 so far has had some of the least volatility. The average monthly risk-free rate shown in Column (3) by far trails this dataset, with at most times the average monthly risk-free rate being essentially zero. It has been worthwhile holding the Thomson Reuters US Total Return Index over the last 7 years, with only once the index underperforming the risk-free rate.

4.3 Is the Strategy Valid in another OECD country that of Canada ? (Research Question 3)

Our original research question states the following: Is the long/short strategy of Boehmer et al. (2010) still valid in an international OECD market that of Canada? OECD stands for the Organisation for Economic Co-operation and Development and consists of 36-member states[52] that are driven to stimulate economic progress and world trade. OECD countries are committed to an open free market and democracy and most members are high income economies with a high human development index.

Employing a strategy like the long/short buying strategy employed by Boehmer et al. (2010) requires an open free market and this is one of the key constituents of an OECD country. The research conducted by Desai et al. (2002), Asquith et al. (2005) and Boehmer et al. (2010) all focused on the American markets. This

[52] The 36 countries in the OECD as of 2019 are as follows: Australia, Austria, Belgium. Canada, Chile. Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Republic of Ireland, Israel, Italy, Japan, South Korea, Latvia, Lithuania, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom and United States.

question gives me the opportunity to move outside of the United States and see if this long/short strategy proposed by Boehmer et al. (2010) is valid in other countries, in particular OECD countries. It will be interesting to see what similarities hold with the US economy and whether there is indeed good news in short interest outside of the United States. By checking whether this strategy is valid for another OECD country like Canada, we will be able to see if the results in the study of Asquith et al. (2005), Boehmer et al. (2010) and Desai et al. (2002) are US specific or apply on an international scale as well. This is of particular interest as strategies employed by regulators and hedge funds in one country may not be applicable in another. We can see the depth of the Boehmer et al. (2010) strategy in another smaller market and whether there are any significant differences between Canada and the US markets. Canada in particular has a strong banking system as seen in Bordo et al. (2015) where Canadian banking stocks were little affected by the 2007-2009 financial crisis, as opposed to banking stocks in other similar OECD countries like the United States, United Kingdom, Germany, France and Italy. Multiple bankruptcies were also recorded in the United States such as Lehman Brothers, Fannie Mae and Freddie Mac. This means there is great investor preference for its banks such as Toronto Dominion Bank, Royal Bank of Canada and Bank of Nova Scotia, this is of interest to us as these investors would be positioning in Canadian banking stocks for their resilience.

It is not possible to focus on all 36 OECD countries due to particular data constraints on short interest limiting my research. However, the Toronto Stock Exchange does publish monthly short interest data, so I can replicate the study in research question 2 for Canada. The study of Canada will constitute my "some international evidence".

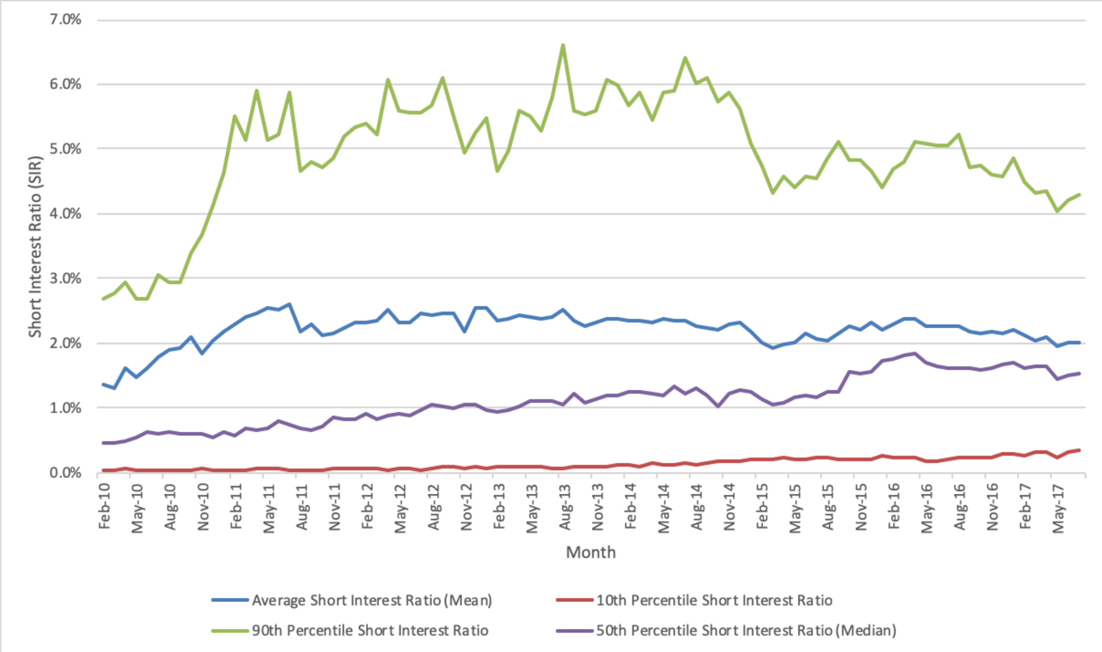
For Canada I use the same time period of the study in research question 2, that of February 2010 to July 2017 in monthly form as to make a like to like comparison with the results of research question 2. I also employ the same Fama and French Three Factor Model with Momentum to adjust for risk premium in monthly form, with again the factor data coming from the Kenneth French Data Library.

The stock returns data for Canada come from the Thomson Reuters Canada Total Return Index and aim to match the Canadian economy as much as possible. Stocks are generally listed on the Toronto Stock Exchange and are most representative of the Canadian economy. The short interest data is from Thomson Reuters DataStream and is again in monthly fashion.

4.3.1 Dataset Statistics for Canada

My dataset consists of 90 monthly observations for the Thomson Reuters Canada Total Return Index between the months of February 2010 and July 2017 inclusive. The Thomson Reuters Canada Total Return Index attempts to model to return of the Canadian economy and consists of stocks mostly listed on the Toronto Stock Exchange. On average, I have around 280 firms per month in each of the monthly observations. I am particularly interested in the short interest that our dataset encompasses as a whole, the level at which the short interest changes each month for the entire dataset and the dispersion of short interest between securities in the dataset.

Figure 6: Distribution of the Short Interest Ratio in the Thomson Reuters Canada Total Return Index



Notes: Data Source using Microsoft Excel from Thomson Reuters Canada Total Return Index DataStream Data. Figure 6 shows the distribution of the short interest ratio in the Thomson Reuters Canada Total Return Index from February 2010 to July 2017. The short interest ratio is calculated as the number of shares sold short divided by the total number of shares in the open market. The Thomson Reuters Canada Total Return Index encompasses my whole dataset.

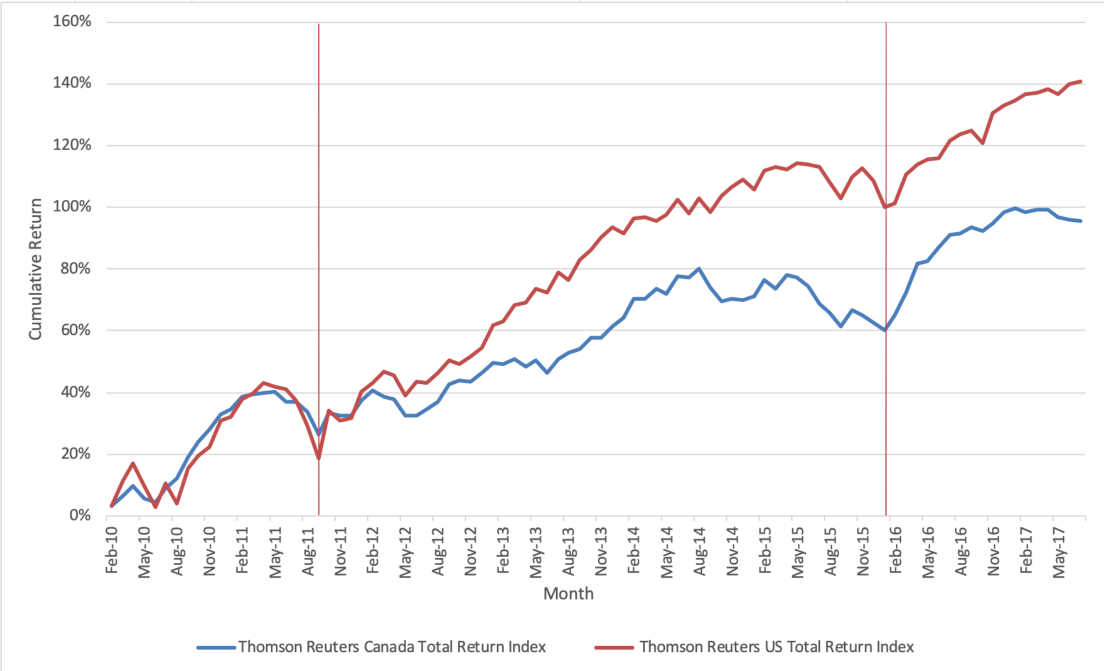
Figure 6 shows the distribution of the short interest ratio in the Thomson Reuters Canada Total Return Index. Looking at Figure 6 we see that the short interest ratio is relatively stable across my dataset, the average short interest ratio (mean) is shown with the blue line in Figure 6 and is calculated by taking the average of all the short interest ratios of each security for a particular month. Each month securities enter and leave the Thomson Reuters Canada Total Return Index and thus the short interest ratio for the index will change each month.

The average short interest ratio is around the 2% mark which is drastically lower than the average short interest ratio we found in the US markets around the 5% mark. This may very well be due to our smaller dataset and generally smaller volumes encountered as a whole in the Canadian markets as opposed to the US markets. It is however a reasonably normal short interest ratio average as US markets in the past have averaged 2% and it is not extraordinary to find on average 2% of a company's total shares floated to be held short.

The other three lines in Figure 6 represent the 90th, 10th and 50th percentile short interest ratios, the percentiles at which those respective percentage of stocks have short interest ratios below. I place emphasis again on the 50th percentile as this represents the median short interest ratio. We can see that again the mean is above the median, showing us negative skew of the data, i.e. there are more stocks with low short interest ratios than there are with high short interest ratios. The negative skew lessens as we move towards 2017, but it is still pronounced.

Looking at the extreme percentiles of 90th and 10th, we see that my dataset exhibits short interest around the 0.2% mark at the 10th percentile level, significantly lower than what we found in the US markets around the 1% mark. The lowest 10% of shorted Canadian stocks have lower short interest than the lowest 10% of US stocks. The 90th percentile is different as Canadian stocks rarely have short interest levels above 6%, while US stocks the 90th percentile is much higher around the 15% mark. Short sellers focus more on the US markets as a whole compared to the Canadian markets judging from this data. The skew between the two countries remains the same however as there are more lightly shorted stocks than heavily shorted stocks in both datasets.

Figure 7: Cumulative Return of the Thomson Reuters Canada Total Return Index VS Thomson Reuters US Total Return Index



Notes: Data Source using Microsoft Excel from Thomson Reuters Canada Total Return Index DataStream Data and Thomson Reuters US Total Return Index DataStream Data. Figure 7 shows the cumulative return of the Thomson Reuters Canada Total Return Index VS Thomson Reuters US Total Return Index. The Thomson Reuters Canada Total Return Index being my dataset and the Thomson Reuters US Total Return Index being the dataset of research question 2. Two vertical red lines indicate the 2011 and 2015 market crisis' which lead to a fall in the total return of the index.

Figure 7 shows the cumulative return of Thomson Reuters Canada Total Return Index vs Thomson Reuters US Total Return Index. Looking at Figure 7 we see that my Canadian dataset does still indeed perform well as a whole, so holding the Canadian market portfolio would have yielded a positive return over those 90 months. There is a great deal of similarity in returns between the two markets until the end of 2011, where then the outperformance of the US markets over the Canadian markets is evident.

Both markets experience dips in 2011 and 2015 (marked with red vertical lines respectively), due to the Eurozone Banking Crisis and the China Yuan Devaluation Crisis respectively, but quickly recover as conditions stabilise. The Eurozone and China effects are global of nature, so it is naturally a contraction in either of those markets would have an effect on all markets in the world, some more than others depending on their exposure to those countries. As time as gone by, world trade has become more and more interconnected and deteriorations in one large global economy are felt around the world

The Canadian dataset is across a complete bull market, as can be shown by the cumulative return in the dataset, it has to be noted that this will affect portfolio performances, as a true bear market would most likely affect all portfolios from the SIR 5% to the SIR 95%. Market risk downturn is more evident in a bear market, where a bull market can produce firm specific risk downturn. In total my Canadian dataset returns 95.6% for the 90-month holding period, as a total return.

Table 8: Distribution of Return in the Thomson Reuters Canada Total Return Index by Year

	(1)	(2)	(3)
Year	Average Monthly Return	Average Monthly Standard Deviation	Average Monthly Risk-Free Rate
2017	-0.4%	1.3%	0.1%
2016	3%	3.4%	0%
2015	-0.6%	3.9%	0%
2014	0.7%	3.6%	0%
2013	1.3%	2.6%	0%
2012	1.2%	3.1%	0%
2011	0%	3.7%	0%
2010	3%	3.1%	0%

Notes: Data Source using Microsoft Excel from Thomson Reuters Canada Total Return Index DataStream Data. Table 8 shows more specifically where the returns are located in the dataset. Data author's own calculation.

Table 8 shows the distribution of return in the Thomson Reuters Canada Total Return Index by year. Looking at Table 8 in Column (1) we can see that the monthly returns vary between -0.6% and 3% on average, with 2015 being the worst year for the dataset and 2016 and 2010 being the best years for the dataset. The standard deviation in Column (2), which is a measure of the variation of returns in my case, shows that 2011 and 2015 were some of the most volatile months, while 2017 so far has had some of the least volatility. The average monthly risk-free rate shown in Column (3) has struggled to be marginally above zero.

Table 9: Standard Deviation in the Thomson Reuters Canada Total Return Index and Thomson Reuters US Total Return Index by Year

	(1)	(2)
Year	US Average Monthly Standard Deviation	Canadian Average Monthly Standard Deviation
2017	1.5%	1.3%
2016	5.1%	3.4%
2015	4%	3.9%
2014	3.6%	3.6%
2013	3.1%	2.6%
2012	3.8%	3.1%
2011	6.6%	3.7%
2010	6.7%	3.1%

Notes: Data Source using Microsoft Excel from Thomson Reuters Canada Total Return Index DataStream Data and Thomson Reuters US Total Return Index DataStream Data. Table 9 shows the standard deviations of the two datasets. Data author's own calculation.

Table 9 shows the standard deviation in the Thomson Reuters Canada Total Return Index and Thomson Reuters US Total Return Index by year. Looking at Table 9 in Columns (1) and (2) we can see that in every case the US dataset is much more volatile than the Canadian dataset. Other than 2014, where both datasets exhibit the same volatility, in every other year the US dataset is more volatile. In the earlier months of 2010 and 2011 the volatility of the US dataset is roughly twice that of the Canadian dataset. The US market volatility may stem from the generally larger short seller base that becomes much more active during market downturns. My assumption is backed up by Diether, Lee and Werner (2009) who show that short selling accounts for one quarter of all trading in the US market. With the large degree of short sellers' present, downward pressure will be on the market in times of market crisis. However, this also means that upward pressure can be exaggerated when short sellers are forced to close positions due to rapidly increasing prices (often referred to as a short squeeze).

4.4 Whether and to what extent short sales affect liquidity, price discovery, volatility and cross-section of stock returns? (Research Question 4)

To answer a question such as this needs a situation where short selling has been banned, whether that is through financial regulatory intervention or a market breakdown. With the ever-increasing size of financial markets and the robust infrastructure, a market breakdown in which short selling does not occur but normal trading does is unlikely. Therefore, my best ability to investigate this question comes through financial regulatory intervention such as that from the SEC or the FCA in regards to short selling.

My original question was the following: Whether and to what extent short sales affect liquidity, price discovery, volatility and cross-section of stock returns? To answer a question such as this I will split my work into three sections, one dealing with the liquidity aspect, the second dealing with the volatility aspect and the third dealing with the price discovery aspect.

The 2007-2009 financial crisis brought a great deal of pain to investors, financial institutions and workforce age people around the world. The downturn of the financial crisis has often been compared to as the second worst economic collapse of the modern world economy behind the great depression of the 1930s in the United States. What was astounding was the V shape recovery of the 2007 to 2009 financial crisis over a prolonged period of mass unemployment and poor growth. The US Federal Reserve, SEC and other foreign security regularity bodies have often been seen as saving the world from the brink of economic depression. The financial system had gone into a period of crisis and the regulations implemented to pull it out from that would be vital for the world. Had regulations not been implemented, a prolonged depression would have ensued. This would have been the norm before the creation of a central bank such as the US Federal Reserve. The effects of the meltdown were still felt 10 years after and it would be a slow and gradual process of normalising central bank policy and central bank balance sheet.

I look to explore the short selling ban imposed by the FCA (formerly known as the FSA in 2008) in financial stocks on the London Stock Exchange Main Market during the 2007-2009 financial crisis. I believe that a great deal of study has been put into exploring the US short sale ban such as Lobanova et al. (2010) and Tian (2014), but very few papers cover the UK short sale ban. As the London Stock Exchange is the third biggest stock market by market capitalisation, I believe a need that this market should be explored in depth regarding short sales.

The FCA banned the short sale of 29 financial stocks[53] between 19th September 2008 and 16th January 2009 in response to the great deal of short pressure put on financial stocks. This was intended to stabilise prices and ease off negative pressure in financial stocks. I look to explore the effect of this short sale ban on the liquidity, volatility and price discovery of the largest of these stocks based on market capitalisation. I produce a sample of the largest 12 banned stocks by market capitalisation and a sample of the largest 12 unbanned stocks by market capitalisation on the FTSE 100. The unbanned stocks act as my control and show that something else isn't causing the change in liquidity, volatility and price discovery.

It is difficult to control the industry effect around the ban, since financial stocks were subject to a short sale ban. This is worth considering when evaluating the results that the Unbanned Portfolio (control) may have trouble in being a true control. However, this doesn't take away the fact that the results from my Banned Portfolio can be evaluated and compared to the best control that we do have.

My dataset consists of two sets of data for both banned and unbanned stocks. I take the top 12 banned financial stocks by market capitalisation and take the top 12 unbanned stocks by market capitalisation. The reason market capitalisation is used is to get the broadest exposure to the market possible, without having an infinite sample of stocks, of which smaller stocks can be missing data. All stocks were constituents of the FTSE 100 in 2008, and many are still. The 12 banned stocks consist of the following companies: Admiral Group, Aviva, Barclays, HSBC Holdings, Lloyds Banking Group, Legal and General, Prudential, Standard Chartered, St James' Place, Standard Life Aberdeen, RSA Insurance Group and the Royal Bank of Scotland

53 Full list of 29 Financial Stocks: Admiral Group, Alliance and Leicester, Alliance Trust, Arbuthnot Banking Group, Aviva, Barclays, Bradford and Bingley, Brit Insurance Holdings, Chesnara, European Islamic Investment Bank, Friends Provident, HBOS, Highway Insurance Group, HSBC Holdings, Islamic Bank of Britain, Just Retirement Holdings, Legal and General, Llloyds Banking Group, London Scottish Bank, Novae Group, Old Mutual, Prudential, Resolution, Royal Bank of Scotland Group, RSA Insurance Group, St James's Place, Standard Chartered, Standard Life and Tawa.

Group. The FCA banned the short sale of these stocks between 19th September 2008 and 16th January 2009. The 12 unbanned stocks consist of the following companies: AstraZeneca, BHP Billiton, BP, British American Tobacco, Diageo, GlaxoSmithKline, Shire, Reckitt Benckiser, Rio Tinto, Royal Dutch Shell, Vodafone Group and Unilever. The FCA did not impose a short sale ban on these companies, though most of these companies (as the general market) struggled during the 2007-2009 financial crisis. Short sell pressure was evident across the market, but focused on financial stocks since the problems had been caused by subprime lending, of which a lot of financial institutions had exposure to.

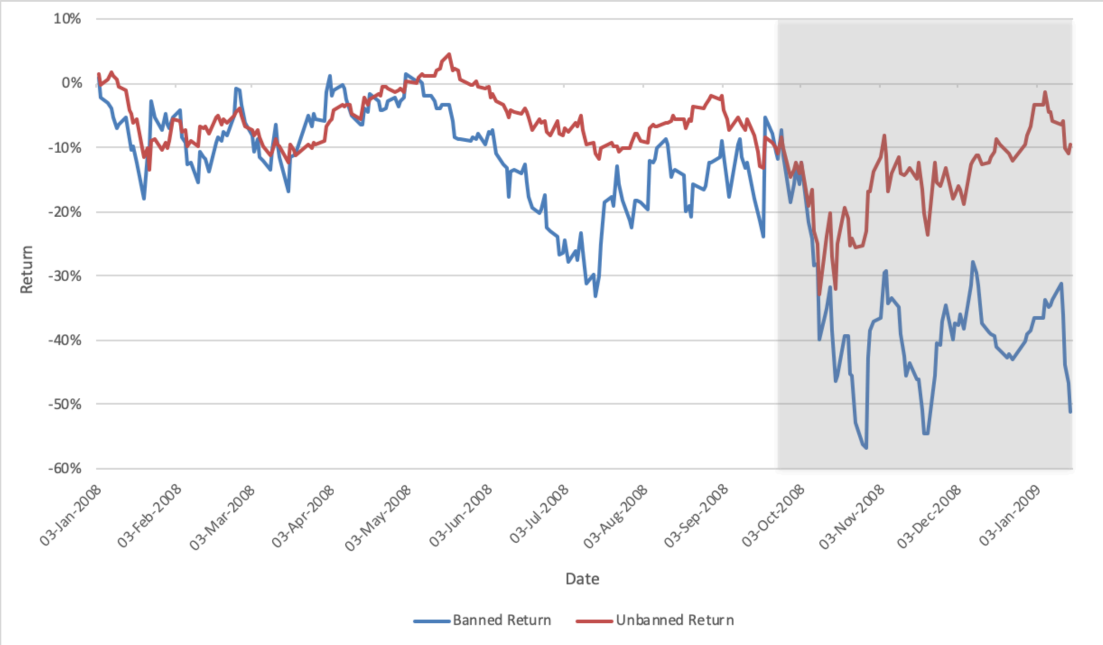
My study period is from 3rd January 2008 to 16th January 2009, which consists of two periods. The first period of 3rd January 2008 to 18th September 2008, in which no stocks have short sale bans implicated by the FCA. The second period of 19th September 2008 to 16th January 2009, where the FCA implicates ban on certain financial stocks to reduce downward pressure on these securities. My study consists of 264 daily observations. I look at the difference between these two periods with respect to volatility, liquidity and price discovery. I collect data for all 24 companies on trading volume, high price, low price and percentage return on a daily basis from 3rd January 2008 to 16th January 2009 using Thomson Reuters DataStream. I also collect trading volume, high price, low price and percentage return for the FTSE 100 as a whole, which is used to represent my market return. This is again using Thomson Reuters DataStream.

4.4.1 Dataset Statistics for the United Kingdom

I am particularly interested in returns, trading volumes and runs distribution of my dataset with respect to both banned and unbanned stocks. Generally, it is expected we will see negative returns for this time period for all stocks in general, higher than average trading volumes and runs with a normal distribution.

Returns are used to in particular for volatility modelling, as changes in returns allow volatility to be seen and fit into respective models such as GARCH. Trading volumes are used in particular in regression models that influence the Bid-Ask Spread. Runs distributions, which are consecutive daily sequences of positive and negative returns are often used to gauge for fat tails and thus price discovery issues in a particular security or portfolio. Fat tails being higher runs distributions in the tails of the distribution than would otherwise be observed if returns were modelled with a normal distribution.

Figure 8: Returns Between Banned and Unbanned Portfolios



Notes: Figure 8 shows the returns between Banned and Unbanned Portfolios. Banned Portfolios consist of stocks banned by the FCA during September 2008 and January 2009. Unbanned Portfolios consist of stocks not banned by the FCA and act as a control. Returns are calculated on a cumulative basis. Grey area on chart indicates short sale ban period.

Figure 8 shows the returns between Banned and Unbanned Portfolios. Highlighted in grey we see the short sale ban period of 19th September 2008 to 16th January 2009. Looking at Figure 8 we see a negative return for both portfolios, which is to be expected since the time period we are exploring is in the depth of the 2007-2009 financial crisis. The Banned Portfolio underperforms the Unbanned Portfolio, this is to be expected since the Banned Portfolio consists of only financial stocks while the Unbanned Portfolio consists of stocks across a vast array of sectors from consumer staples, pharmaceuticals, oil, alcohol, tobacco and telecommunications. Financial stocks were hit the most since the crisis stemmed from subprime lending and companies with exposure to this felt the worst effects. Certain sectors are seen to be termed defensive, such as alcohol and tobacco, where economic conditions generally do not seem to affect earnings per share[54]. Financials are seen to be economy sensitive and do underperform in recessions, in my case the recession stemmed out of the financial sector, which in turn put further pressure on these financial stocks.

For the period before the short sale ban, the Unbanned Portfolio averages -0.1% return in a day, during the short sale ban the unbanned portfolio averages 0% return in a day. For the period before the short sale ban, the Banned Portfolio averages -0.1% return in a day, during the short sale ban the Banned Portfolio averages -0.3% return in a day. The underperformance of banned stocks started before the short sale and once the

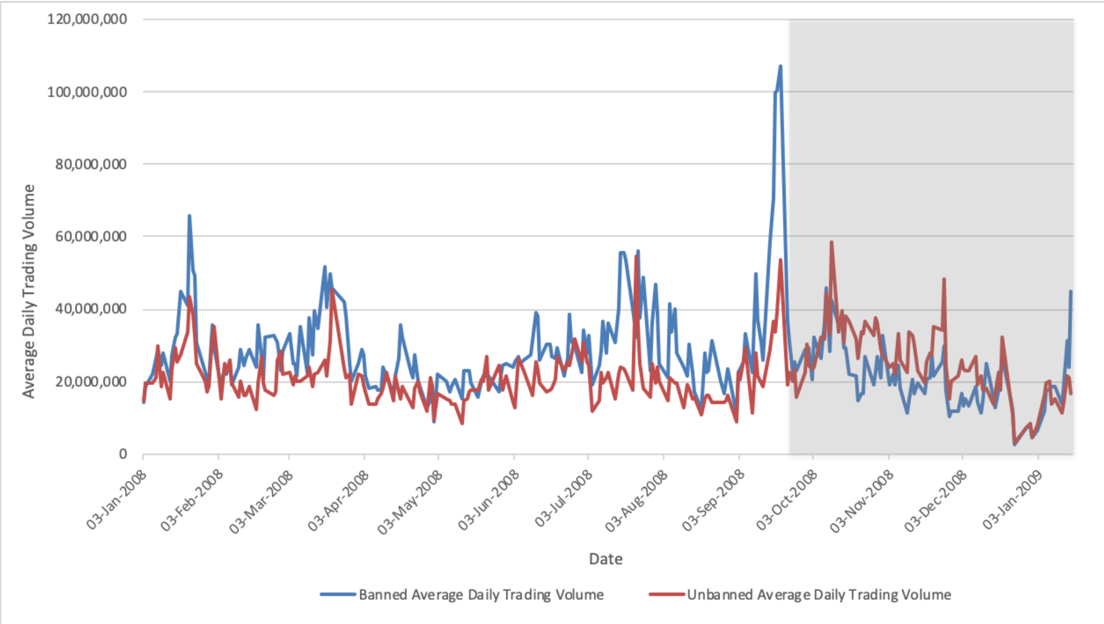
[54] If earnings per share are not affected, a stock price is unlikely to drop (as long as forward guidance is maintained). The stock price is a combination of the earnings per share assigned a market multiple to reflect future perceived earnings growth. For example, a company growing earnings at 10% a year and earning £1 a share currently, should be trading around £10 to be fairly priced. However, supply and demand from investors and in particular speculators will mean a market price will often move away from fair valuation.

short sale ban takes place on 19th September 2008, we see even more divergence in performance between portfolios. It looks like the short sale ban does very little to curb the selling in these banned stocks.

The sharp drop in September is attributed to the collapse of Lehman Brothers, an investment bank in the United States which had large exposures to subprime loans. The banned stocks also seem to lack market direction and show exaggerated swings, this is where short sellers could help in finding a top and bottom. What is evident though is that there is a lot of risk aversion with the holders of banned stocks, as going to zero in stock price is most definitely an option with the collapse of AIG, Lehman Brothers, Bear Sterns, Freddie Mac and Fannie Mae; the governments are willing to let firms go bust. With a market taken over by risk aversion, everyone is trying to rush for the exit, the asymmetric nature of stock returns is evident here with sharp falls, those confident enough to buy during these times were rewarded very well a few years out.

If the regulatory authorities had saved Lehman Brothers, these falls in other financial stocks I believe would not be as steep as a lower bound above zero would be established at least in terms of near book value at an absolute minimum. Also, the formation of the Troubled Asset Relief Program[55] (TARP) after the rescue of Lehman Brothers would have further assured the market. However, the complacency of the management of Wall Street Firms would be much higher today if they were aware of a "too big to fail" culture and it is very possible that the same aggressive risk taking would be present today.

Figure 9: Trading Volume Between Banned and Unbanned Portfolios



Notes: Figure 9 shows the average daily trading volume between Banned and Unbanned Portfolios. Banned Portfolio consist of stocks banned by the FCA during September 2008 and January 2009. Unbanned Portfolios consist of stocks not banned by the FCA and act as a control. Returns are calculated on a cumulative basis. Grey area on chart indicates short sale ban period.

55 Troubled Asset Relief Programme (TARP) was a programme instigated by the United States government in order to strengthen the financial sector of the United States. The programme involved the purchase of troubled assets from US financial institutions, thus removing these assets from the financial system and injecting cash from US taxpayers.

Figure 9 shows the average daily trading volume between Banned and Unbanned Portfolios. Highlighted in grey we see the short sale ban period of 19^{th} September 2008 to 16^{th} January 2009. Looking at Figure 9 we see that the trading volume between Banned and Unbanned Portfolios has a good level of correlation. The correlation coefficient[56] for the pre-short sale ban period is 0.673 and the correlation coefficient for the short sale ban period is 0.623. So, we see a decrease in correlation between trading volumes, once the short sale ban takes place.

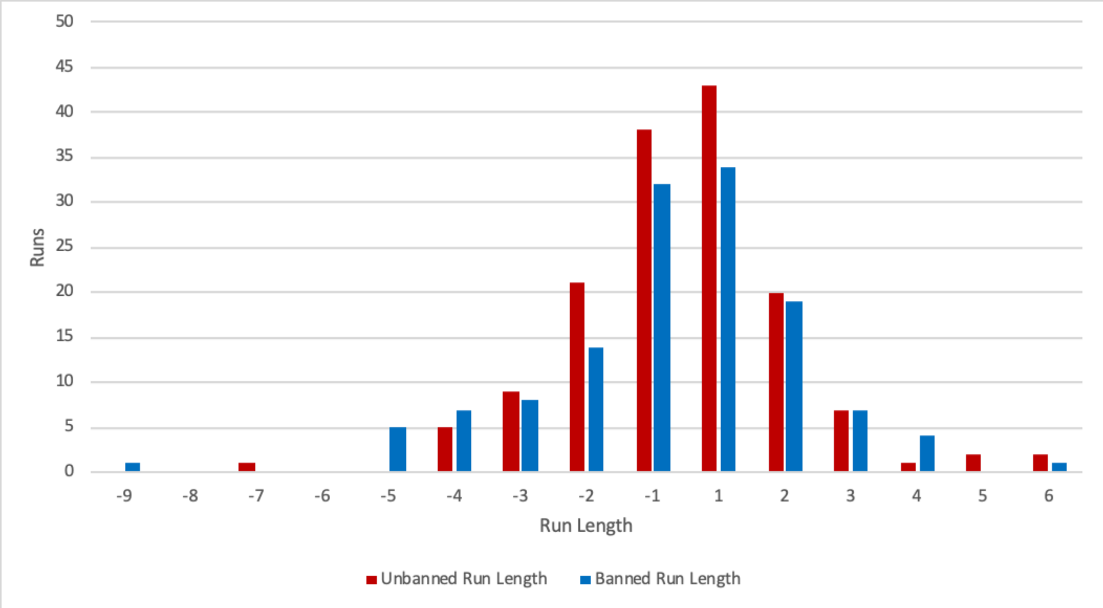
The big spike in banned trading stocks volume in September 2008 is attributed to the collapse of Lehmann Brothers, where inflows of volume would have been seen by investors. What is noticeable is the trading volume of Unbanned Portfolios is generally lower until the short sale trading ban is imposed on September 2008. Then we see a lowering of trading volume in the Banned Portfolios, showing the absence of short sellers.

For the period before the short sale ban, the Unbanned Portfolio averages 20,980,652 shares traded in a day, during the short sale ban the Unbanned Portfolio averages 25,072,496 shares traded in a day. For the period before the short sale ban, the Banned Portfolio averages 29,360,758 shares traded in a day, during the short sale ban the Banned Portfolio averages 22,236,268 shares traded in a day. Clearly the short sale ban has affected trading volumes.

Trading volumes for both portfolios are seen to be with no clear uptrend or downtrend, which is to be expected over a period of a year. Over a longer period of decades, we see trading volumes have dramatically increased with the ease of access to markets that were not there in the past. There is also a monthly dependent nature to trading volumes that needs to be observed, August is usually quiet when it comes to volumes with most investors and traders on vacation, volumes are higher in September where these people return. A lot of trading is done algorithmically as well, meaning volumes are generally higher than a few decades ago just from computers trading with other computers.

[56] The correlation coefficient is used to measure the correlation between two sets of data and takes the range between -1 and 1, where -1 shows perfect negative correlation, 1 shows perfect positive correlation and 0 shows little to no correlation.

Figure 10: Banned Portfolio and Unbanned Portfolio Runs Distribution



Notes: Figure 10 shows the distribution of runs for the Unbanned Portfolio and Banned Portfolio for the entire length of the dataset. A run constitutes a consecutive sequence of either positive or negative daily returns for each portfolio. A zero-percentage daily return continues the run. A run only ends where there is a sign change in the daily return. Positive run lengths indicate positive runs, and negative run lengths indicate negative runs.

Figure 10 shows the runs distribution between Banned and Unbanned Portfolios. A run is categorised as a period of daily returns which are of the same sign. Once the sign of the daily return changes, the old run ends and a new run starts. A zero-percentage daily return continues the old run. The length of a run is the number of consecutive daily returns of the same sign (either positive or negative). It is worth mentioning we do not define a run of length 0, as a run of length 0 would not be possible as a run is either positive or negative of which zero is neither.

Runs are a technique that are often used to calculate the effects of short selling on price discovery and are of interest in this thesis. This is because a runs distribution allows a researcher to explore both positive and negative fat tails, which are often a cause of short selling impeding the price discovery process.

We can initially see across the entire dataset, both portfolios of runs form a normal distribution, with more runs focused on the centre of run length 1 and -1, as opposed to runs focused on the extremes of length 6 and -9. This is to be expected as market movements often move either up or down in a zig zag fashion with retracements, it is very rare to see a consistent string of up moves or down moves over a 10 day plus period.

The Unbanned Portfolio has a higher proportion of runs near the centre compared to the Banned Portfolio showing that the Unbanned Portfolio is of a much less volatile nature. The majority of runs of both portfolios are of a negative nature showing a negative return across the period, which can be seen in more detail with the returns data.

Looking at the tails of the distributions we see the Banned Portfolio exhibits the fatter tails with a run of -9 and a run of 6, as opposed to the Unbanned Portfolio which has a tail with a run of -7 and a run of 6. Overall it can be seen that the Banned Portfolio is the much riskier to hold and a more benefit to speculators over investors over this study period.

4.5 Conclusion

In conclusion we have seen the data description for my research questions and the dataset statistics that make up my datasets. It is very important that a good analysis of the data is given as it allows the researcher and reader to see reasons for certain results. I provide a great deal of dataset statistics for research questions 1, 2 and 3 in the form of the short interest ratio distribution and the return of the dataset as a whole. This is a great comparison when looking at individual portfolio returns and the short interest levels that are generating these returns. Short interest ratio levels have dramatically changed in my dataset and in the preceding decades, so quantifying high or low short interest ratios in very meaningful.

My dataset statistics are consistent with past studies regarding levels of short interest ratio and general returns of the market around 10% a year with the inclusion of dividend reinvestment. Short interest ratio levels, although skewed to the lower percentiles, fit a normal distribution with skew. My dataset does not show any evidence of a high concentration of stocks with very high or very low short interest ratio levels. Returns look consistent with the market as a whole, and I do not see any periods of sustained high returns or sustained low returns. We see market crashes, but quick recoveries meaning we are reverting to the mean return of 10% a year with the inclusion of dividend reinvestment. Reversion to the mean is an important concept which states that the underperformance or overperformance of securities is not a long-term affair and stocks will revert to their mean return in the long run. This is natural as greed will often lead to the overpricing of securities and fear will lead to the under-pricing of securities, however we revert to the mean as greedy buyers are exhausted and fearful sellers are exhausted. This naturally leads to a top in stocks and a bottom in stocks. In my case I believe that this indeed does hold true.

With research question 4 we see the data description in regards to returns, trading volume and runs in a portfolio of stocks which has not experienced a short sale ban and a portfolio of stocks which has experienced a short sale ban. These three metrics are essential to form the methodologies on investigating volatility, liquidity and price discovery.

We see returns as expected in terms of the Banned Portfolio underperforming the Unbanned Portfolio, we see a decrease in trading volume as expected when a short sale ban is enforced on the Banned Portfolio. We also see a wider runs distribution in the Banned Portfolio, again showing the wider volatility profile the Banned Portfolio holds over the Unbanned Portfolio. I again see dataset statistics that are consistent with past studies.

CHAPTER 5: ESTIMATION PROCEDURES

The estimation procedures chapter is concerned with the methodology I am employing to my data to investigate my hypotheses. It is crucial that we use appropriate estimation procedures and make it clear to the audience why we are using them as they will be essential in the validity and replicability of my results. The estimation procedures chapter like the data description chapter, has been split up into the respective research questions. Estimation procedures in section 5.1 report for research question 1, estimation procedures in section 5.2 report for research question 2, estimation procedures in section 5.3 report for research question 3 and estimation procedures in section 5.4 report for research question 4.

5.1 Five Factor Model as an Adjuster for Risk Premium (Research Question 1)

The methodology for this research question is in line with Boehmer et al. (2010), with the main difference being the model that is used to adjust for risk premium. Boehmer et al. (2010) used the Carhart Four Factor Model to adjust for risk premium, while my study will involve using the more advanced Fama and French Five Factor Model.

On a historical basis, short interest data has been published on the Wall Street Journal on a monthly basis before September 2007 and on a bi-monthly basis after September 2007. I will stay consistent with Boehmer et al. (2010) and use short interest data on a monthly basis to form my portfolios each month and rebalance every month as necessary. My short interest data provider will be Thomson Reuters DataStream that will provide the short interest data for each security each month.

To begin my analyses, I will form two highly shorted and two lightly shorted portfolios each month. The highly shorted portfolios will include stocks from the 95^{th} and 90^{th} percentile of the short interest ratio distribution in the previous month based on the data from Thomson Reuters DataStream. The lightly shorted portfolios will include stocks from the 5^{th} and 10^{th} percentile of the short interest ratio distribution in the previous month based on the same data from Thomson Reuters DataStream. Stocks will enter and leave the portfolios depending on their previous month level of short interest, this portfolio construction technique is often referred to as Calendar Time Portfolios and is employed similarly in Desai et al. (2002) and Asquith et al. (2005). The nature of the rolling portfolios means it is unlikely two portfolios will ever be the same each month, this adds a dynamic nature to my study.

I will also be constructing long/short portfolios to test for significant differences between the 95^{th} and 5^{th} percentile portfolios and 90^{th} and 10^{th} percentile portfolios. The Fama and French Five Factor Model will be used as an adjuster for risk premium for the excess monthly return of these six portfolios, the four standard portfolios and the two long/short portfolios.

In regards to persistence in the portfolios the 95th Percentile Portfolio has a mean number of stocks in the portfolio as 64, the 90th Percentile Portfolio has a mean number of stocks in the portfolio as 127, the 5th Percentile Portfolio has a mean number of stocks in the portfolio as 64 and the 10th Percentile Portfolio has a mean number of stocks in the portfolio as 127.

The maximum number of stocks in the 95th Percentile Portfolio is 80 and the minimum number of stocks in the 95th Percentile Portfolio is 47. The maximum number of stocks in the 90th Percentile Portfolio is 160 and the minimum number of stocks in the 90th Percentile Portfolio is 94. The maximum number of stocks in the 5th Percentile Portfolio is 80 and the minimum number of stocks in the 5th Percentile Portfolio is 47. The maximum number of stocks in the 10th Percentile Portfolio is 160 and the minimum number of stocks in the 10th Percentile Portfolio is 54.

There is variation as expected in the number of stocks in each portfolio, this is due to stocks either delisting off the exchange or entering a portfolio after the 1-year IPO exclusion period. Also, there may be cases where there is no data available for a stock in terms of either returns or short interest affecting its inclusion in a portfolio.

The types of companies found in the highly shorted portfolios (95th and 90th percentile) usually include small cap companies with poor financial health, lack of material earnings, and low trading volumes. These are often referred to as penny stocks. In the lightly shorted portfolios (5th and 10th percentile) we usually find large or mid cap companies with good financial health, good material earnings and high trading volumes. These are often referred to as investment grade stocks. This is not always concrete and there can be large cap stocks in highly shorted portfolios and small cap stocks in lightly shorted portfolios.

Boehmer et al. (2010) in their study constructed three highly shorted (99th, 95th and 90th percentile), three lightly shorted (1st, 5th and 10th percentile) and the respective long/short portfolios (1st-99th, 5th-95th and 10th-90th percentile). However, from some pilot testing I am against constructing portfolios for the 1st and the 99th percentile of stocks, I believe that in particular for smaller datasets[57] these portfolios hold far too few stocks that firm specific characteristics are more likely to be exaggerated in the results than short interest specific characteristics. As the number of stocks in my portfolios increase, the more so the result of short interest is reliable and the less I have to fear of firm specific characteristics influencing my results.

In cases of smaller datasets there may be cases of only 3 or 4 stocks being included in the 1st and 99th percentile portfolios and this I believe is a risk to the integrity of my results. By holding 3 or 4 stocks the diversifiable risk of the portfolio will not be removed and will also pose a volatility issue for the market participants holding such a portfolio.

[57] A sample size that is too small leads to reduction in power of the study and increases the margin of error, which can render the study meaningless. The power of the study is of its ability to avoid Type II errors. A Type II error occurs when the results confirm the hypothesis on which the study was based when in fact an alternative hypothesis is true. Using a sample size that is too small increases the likelihood of a Type II error skewing the results. This leads to a decrease in power of the study.

$$E(R_i) - R_f = \alpha_i + \beta_i\big(E(R_m) - R_f\big) + \gamma_i E(SMB) + \delta_i E(HML) + \eta_i E(RMW) + \lambda_i E(CMA) + \varepsilon_i \quad (36)$$

The Fama and French Five Factor Model as presented in equation (36) is used as the adjuster for risk premium. The model reports the expected return of a security or portfolio using factors that in the past have contributed significantly to security returns. $E(R_i) - R_f$ is the excess return of each individual portfolio above the risk-free rate which I will be using as the US 3-month treasury bill rate.[58] The coefficient α_i is a very significant as it will show how well the Fama and French Five Factor Model matches returns on my portfolios. An α_i of zero indicates that the model explains all returns, while a significantly high or negative α_i indicates that the model shows an under or overperformance of returns. β_i is the beta of my portfolios and will indicate how volatile they are in respect to the market. $E(R_m) - R_f$ indicates the equity risk premium my portfolios hold, and the excess return the market is able to generate above the risk-free rate.

The coefficient γ_i shows how much of my portfolio return is contributed to the SMB factor. The SMB factor states that on average in the past, small firms have outperformed big firms taking all other factors into account. The coefficient δ_i shows how much of my portfolio return is contributed to the HML factor. The HML factor states on average in the past, low book to market firms have outperformed high book to market firms taking all other factors into account. The coefficient η_i shows how much of my portfolio return is contributed to the RMW factor. The RMW factor states on average in the past, taking all other factors into account, high profitability firms have outperformed low profitability firms. The coefficient λ_i shows how much of my portfolio return is contributed to the CMA factor. The CMA factor states on average in the past, taking all other factors into account, high investment firms have outperformed low investment firms. ε_i is the error term and takes into account errors in the regression model, if the regression model is correctly specified according to OLS the ε_i should have a mean equal to zero.

In terms of study objectives of short interest, the Fama and French Five Factor model shows us where in particular the returns are coming from. These are based on the previously mentioned five factors of market risk premium, SMB, HML, RMW and CMA. The α_i will show us the under and overperformance of my model in relation to short interest, with a high α_i, we will see that my model is overperforming based on either high or low short interest portfolios while a low α_i will show that my model is underperforming based on either high or low short interest portfolios. Based on previous literature I expect the underperformance of heavily shorted stock portfolios (shown with a lower α_i than lightly shorted portfolios).

My rationale for using the Fama and French Five Factor model is based on its general acceptance as one of our best asset pricing models in academic finance. There are however studies such as Blitz et al. (2018) which raise valid criticisms against the model, but I am aware that a great quantity of asset pricing models have and still to this day face criticism regarding particular factors chosen. However, models built up from the CAPM

[58] This is the standard in the risk-free rate, many other empirical studies such as Desai et al. (2002), Asquith et al. (2005) and Boehmer et al. (2010) have used this risk-free rate as their default. Currently the shortest-term government debt security of the world's strongest economy (United States) in 2019 is the best standard for the risk-free rate.

are very good at modelling returns, and in addition no model can be perfect given the ever-changing nature of the markets, both in terms of participants and securities traded.

For my six portfolios which change stock composition every month, I run a time series regression [59] on portfolio excess returns using this Fama and French Five Factor Model. I report for equal weighting of portfolios.[60] It has been sometimes put forward that value weighting is better in terms of performance evaluation, because value weights often reflect the holdings of the average investor. The drawback of value weighting is that it won't show the average investor's net short position. The average investor (in contradiction to the average short seller) has a zero-short position at all times. On the long side the supply of shares of a certain stock is seen to be fixed (in the short term at least), but this is not seen to be true for short interest. Short interest can vary month to month particularly for individual stocks. We can also see that if a stock that was heavily shorted had a persistent negative return, in my value weighted portfolio it would receive a smaller weight each month, masking the success of short sellers. Therefore, I will be reporting for equal weighting. Thus, given the above reasons, it has been seen by me and the current literature equal weighting may be superior to value weighting when you are studying the performance of stocks that have been shorted.

The calculation for equal weighting is relatively simple. For example, for a portfolio with n securities, equal weighting can be calculated with equation (37) as follows:

$$Equal\ weighting\ for\ stock\ x\ = \frac{100}{n} \tag{37}$$

Where x is the weight allocated to each security and n is the number of securities in the portfolio.

Given I am running a multiple linear regression model as my central econometric technique, I have to test for assumptions that the multiple linear regression model holds to remain BLUE (best linear unbiased estimator). Violation of the Gauss Markov Theorem means that my estimator will not be BLUE. My multiple linear regression model will use the excess portfolio returns as the independent variable and the sensitivity of the factor loadings as my dependent variables.

Several key assumptions need to be held, the first being multivariate normality. This states that the residuals from the regression need to be normally distributed. We must also have the assumption that there is no multicollinearity, this means the independent variables are not highly correlated with each other. The last

59 It is worth noting that my regression is of a time series nature and not a panel data nature. I am regressing portfolio excess returns against the Fama French Factors that vary in a series of monthly time intervals. Fama-MacBeth regression is a popular technique that is of a panel data nature that can also be implemented to estimate parameters. We would regress each asset against the proposed risk factors to determine the asset's beta for that risk factor, then all asset returns are regressed for a fixed time period against the estimated betas to determine the risk premium.

60 The difference between value and equal weighting is that portfolios which are equal weighted, will have an equal weight of each security in the portfolio. Portfolios which are value weighted will be weighted by market capitalisation, where the larger market cap firms compose more of the portfolio than the smaller market cap firms.

assumption that needs to be satisfied is of homoscedasticity. This assumption states that the variance of the error terms is similar across all values of the independent variables.

I will check for these assumptions using specially designed tests. Normality of residuals can be checked using a goodness of fit test such as the Kolmogorov-Smirnov Test. Multicollinearity can be checked using a Correlation matrix or the Variance Inflation Factor, I will do both. Finally, homoscedasticity can be checked using a scatterplot of residuals versus predicted values. There should be no clear pattern in the distribution and if there is a cone shaped pattern, my data is said to be heteroscedastic. Doing all these tests will ensure the robustness of my regression results and aid in extrapolation if needed.

In this study I am particularly interested in those long/short portfolios as was Boehmer et al. (2010) where they showed that there was indeed good news in short interest, by going long a lightly shorted stock portfolio and going short a heavily shorted stock portfolio and rebalancing each month. It is worth mentioning the nature of short interest and the general distribution we are likely to encounter. The nature of short interest means that as financial bull markets generally rise over time as seen with any long-term S&P 500 graph, we are likely to see more stocks with higher short interest levels and less stocks with lower short interest levels. This is based on the fact that short sellers are efficient in targeting overpricing as seen in Comerton-Forde and Putnins (2009) and Chan et al. (2017). Short interest is based on the change in level rather than the absolute level. Across decades, this will rebalance as markets become oversold in a recession and short sellers retreat by closing positions as they incur losses once a technical bull market takes formation. This will mean that we are not going to have portfolios with equal amounts of stocks in them, a similar situation Boehmer et al. (2010) encountered as well.

Boehmer et al. (2010) found that during the first few years of their study, nearly 20% of stocks had zero or negligible short interest in them. So, the heavy skew towards low short interest is evident in their sample and will most likely be evident in mine since I will be pulling stocks from the same NYSE and Nasdaq exchanges. However, it must be noted that shorting as an activity has dramatically increased over the years and is of course particularly heightened in bear markets.

The introduction of the options markets [61] has also contributed to the shorting activity as seen in Figlewski and Webb (1993). It means less capital can be employed to obtain a larger risk position, which would not have been possible in the past. The markets have also become more retail friendly as trading takes place outside the stock exchange floors,[62] which again has contributed to the rise in shorting activity. More and more CFD (contract for difference) and spread betting firms have been brought into the market, which has made shorting a standard activity, which in the past was reserved for the hedge funds and other institutional managers. Other than significant shorting bans, which are usually on a small and temporary basis, shorting activity and short interest has gradually increased over the decades, this can be seen in Desai et al. (2002),

61 The options market has meant you can take a relatively large short position using put options rather than borrowing and selling the security itself in the standard manner. This amplifies gains and losses, since being on the losing side in an options contract means you lose 100% of the principal when the contract expires. However short selling can incur losses over 100% of invested capital.

62 Trading before the rise of the internet was restricted to exchange floors such as the NYSE, orders would be sent via telephone or made in person on the floor of exchanges. Now exchanges are very quiet with most orders entering the system electronically, increasing speed and efficiency of the market. There is also no delay in prices, as order books show the open bids and offers in live time.

Boehmer et al.(2010) and in this thesis in Chapter 4 of Data Description . For the time period of this dataset there have been no significant shorting bans for United States securities.

It was also noted in the study of Boehmer et al. (2010) that larger stocks were in general less likely to have low short interest compared to small and mid-cap stocks. This may very well be due to the larger volumes that are involved with large cap stocks, while small cap stocks can suffer from liquidity issues at times. If the large institutional investors are restricted from investing in certain small cap stocks, it can cause a loss of volume.

Refer in addition to Appendix 2 for a methodology diagram outlining the estimation procedures implemented to answer research question 1. The methodology diagram allows us to see the step by step processes that need to take place in the method to obtain the results.

5.2 Is there Arbitrage? (Research Question 2)

The methodology for this question will again be in line with Boehmer et al. (2010), however in this case I will be interested in whether the findings of Boehmer et al. (2010) are still robust after the publication of their study. I will be exploring the period after the publication of Boehmer et al. (2010), as that is when this strategy has become known to the market. There are many cases where market arbitrage takes place and makes a once risk-free strategy invalid, given the debate of market efficiency, a perfectly efficient market would remove such strategies once information regarding these strategies became public.

It has clearly been shown by Boehmer et al. (2010) that by simply holding a long/short portfolio based on the highest and lowest short interest ratio and rebalancing every month, one can achieve a risk-free adjusted return. However, I am interested whether hedge funds [63] have implemented this strategy as they in particular have the potential and knowledge in terms of finance and coding software to implement such a strategy with relative ease.

It is worth mentioning that there are several drawbacks to this strategy in terms of the rebalancing costs. An investor would be required every month to sell and buy stocks to maintain a long/short portfolio based on new short interest data, if the investor is dealing with a large number of buy and sell orders, this can lead to a dramatic increase in transaction fees. In the UK for example every buy and sell order by retail investors will usually encounters a fee on top of stamp duty for buying a security outside of the FTSE AIM (alternative investment market). A separate fee would be on the sell side and further taxation charges could occur if investing outside of an ISA (individual savings account). On average a buy or sell order of a security in the UK in September 2019 costs around £7, with some firms charging more or less. It is common knowledge on Wall Street, that by over trading only brokers are made rich, while by simply buying and doing nothing the returns are in the investors' favour. That is often why many brokerage houses on Wall Street are said to merit speculation over investment, since the brokerage houses can get great amounts of commission by the

[63] Hedge funds are institutions that are solely focused on earning a return for investors and thus taking a fee based on money held and/or performance. They are different from investment banks in the sense that they usually do not take part in activities such as market making, corporate restructuring and mergers and acquisitions.

constant stream of buying and selling. This usually involves investors often buying high and selling low based on hope, fear and greed.

The strategy is also very diversified, and this may put off some hedge funds, as the returns being less diversified could be greater. Hedge funds may choose to implement this strategy on a smaller scale and focus on good companies that exhibit low short interest and target the worst companies that exhibit high short interest. It will be interesting to see the position hedge funds and in particular retail investors have taken regarding this strategy. It is almost certain that the monthly selling and buying of securities would cut into retail investors profits, but with sufficient capital employed it may be worthwhile for them as well.

The strategy is however very systematic and could be easily implemented by an algorithm,[64] this does add to the positives of the strategy. The biggest hedge funds would have the resources to implement a strategy such as this algorithmically and this does need to be noted, as the normal retail investor would not have access to such programmes and systems.

I will again form six portfolios to test this strategy. I will form two portfolios on the 5th and 10th percentile on the short interest ratio of the previous month and two portfolios on the 95th and 90th percentile on the short interest ratio of the previous month. The portfolios will change each month depending on the short interest ratio of the previous month, so therefore the number of stocks per month in each portfolio will be different as well. I will also be constructing long/short portfolios to test for significant differences between the 95th and 5th percentile portfolios and 90th and 10th percentile portfolios. Again, I am aware of the limitations of the 1st and 99th percentile portfolios and as similar to research question 1 I exclude these portfolios from my study.

I will then regress the excess return of these portfolios using a time series regression. The model employed will be the Fama and French Three Factor Model with Momentum shown in equation (38), also known in the literature as the Carhart Four Factor Model. This will be my primary model that will adjust for risk premium. This is the same model employed by Boehmer et al. (2010) and Desai et al. (2002) to adjust for risk premium.

$$E(R_i) - R_f = \alpha_i + \beta_i\left[E(R_m) - R_f\right] + \gamma_i E(SMB) + \delta_i E(HML) + \eta_i E(MOM) + \varepsilon_i \tag{38}$$

Where $E(R_i)$ is taken as the expected return on asset *i*. R_f is the risk-free rate of return, taken as the 3-month US treasury rate. β_i is the systematic risk of asset *i*, sometimes known as just the risk of asset *i*. $\left[E(R_m) - R_f\right]$ is the market price of risk, which is the excess return above the risk-free rate that holding the entire market portfolio is able to achieve. γ_i is the factor loading for SMB (Small Minus Big). SMB being the factor that notes the difference between the outperformance of small firms compared to large firms. δ_i

64 A simple algorithm could be set to scan short interest data just published, compare it to the current long and short portfolio held by the hedge fund. Securities which do not fit the criteria in the current portfolio could be removed and securities which do fit the criteria could be added. The portfolio would automatically adjust each month based on new short interest data, making the process seamless and efficient.

is the factor loading for HML (High Minus Low). HML being the factor that notes the difference between the outperformance of growth stocks compared to value stocks. η_i is the factor loading for MOM (Momentum). MOM being the factor that notes the difference between the outperformance of stocks showing short term momentum over stocks without short term momentum. ε_i is the error term to account for inconsistencies in the model. The model is standard in adjusting for risk premium.

I will also be implementing the equal weighting strategy where the composition of stocks in each portfolio will be of equal weight rather than weighted by market capitalisation or any other metric. Most investors outside of exchange traded funds (ETFs) tend to hold portfolios that are equal weighted when selecting equities and this construction adds to mimic their behaviour. The average investor may not been be aware of the market capitalisation of the companies they hold and would have be constantly adjusting weightings based on changes in market capitalisation.

In terms of study objectives of short interest, the Fama and French Three Factor Model with Momentum shows us where in particular the returns are coming from based on the previously mentioned four factors of market risk premium, SMB, HML and MOM. The α_i will show us the under and overperformance of my model in relation to short interest, with a high α_i, we will see that my model is overperforming based on either high or low short interest portfolios while a low α_i will show that my model is underperforming based on either high or low short interest portfolios. Based on previous literature we expect the underperformance of heavily shorted stock portfolios (shown with a lower α_i than lightly shorted portfolios).

Regarding the regression I will again be checking for multivariate normality, no multicollinearity and homoscedasticity in the results. I will perform tests in line with research question 1 to check for these conditions and to amend the dataset if necessary to counteract these conditions. By holding these conditions, I will ensure my model is BLUE.

As far as I am aware there have been no significant shorting bans for United States securities from the period of February 2010 to July 2017. If such a ban existed, I would exclude the monthly short interest data for each security until new short interest data became available, this is in line with Desai et al. (2002) and Boehmer et al. (2010). Therefore, months would be missing from my time series in order to reflect missing data.

Refer in addition to Appendix 3 for a methodology diagram outlining the estimation procedures implemented to answer research question 2. The methodology diagram allows us to see the step by step processes that need to take place in the method to obtain the results.

5.3 Is the Strategy Valid in another OECD country that of Canada ? (Research Question 3)

This question aims to see whether what I have found in research question 2 is applicable across an OECD country that of Canada. The methodology for this question will be in line with Boehmer et al. (2010) however

I will be exploring a market that has never been explored before, that of Canada. This is the novelty in the question. In an ideal world, I would test extensively across the OECD base, however lack of short interest data in a lot of major markets has meant I have had to focus on a market where the short interest data is available, that of Canada.

Even though Canada is listed as an OECD country, there are distinctions between Canada and the United States in terms of security listing numbers and general interest in markets. There is a vast global interest in United States equity markets, as opposed to Canadian markets which often experiences more national interest rather than international interest. This will mean there will be differences in results compared to the United States, which is of interest to me. Canada also has a very stable banking system that is often preferred by investors over many countries, this will also be of interest with the results. Canadian Banking Stocks such as Royal Bank of Canada, Bank of Montreal, Toronto-Dominion Bank and Bank of Nova Scotia are of some of the highest quality in the world both in terms of balance sheet and their dividend pay-outs. There is a great deal of preference for investors in Canadian Banking Stocks who want to balance growth and income.

I will again form six portfolios to test this strategy. I will form two portfolios on the 5^{th} and 10^{th} percentile on the short interest ratio of the previous month and two portfolios on the 95^{th} and 90^{th} percentile on the short interest ratio of the previous month. The portfolios will change depending on previous month short interest and therefore will have a different number of stocks each month. I will be again constructing long/short portfolios to test for significant differences between the 95^{th} and 5^{th} percentile portfolios and 90^{th} and 10^{th} percentile portfolios. I again leave out the 1^{st} and 99^{th} percentile portfolios and especially so since we are working with a much smaller dataset for the Canadian market, I again do not believe 3-4 stocks in a portfolio will be representative of short interest characteristics rather than firm specific characteristics. Boehmer et al. (2010) would have experienced a great deal of volatility with smaller portfolios and I believe that this will not add to my results.

The excess return on each portfolio will be regressed using a time series regression and the model used to adjust for risk premium will be the Fama and French Three Factor Model with Momentum shown in equation (39), also known in the literature as the Carhart Four Factor Model. I chose this model in particular as to make a direct comparison as possible with research question 2, as I have kept everything constant other than the market being studied. The variables which are constant are my control variables and the variable which is changing, that of the Canadian market is my test variable.

The model I am using to adjust for risk premium is the same model implemented in research question 2, Boehmer et al. (2010) and Desai et al. (2002). It is one of the most widely used models to adjust for risk premium in the literature to date, although I believe more and more of the literature will shift towards the Fama and French Five Factor Model once its consistency is shown across different markets.

$$E(R_i) - R_f = \alpha_i + \beta_i[E(R_m) - R_f] + \gamma_i E(SMB) + \delta_i E(HML) + \eta_i E(MOM) + \varepsilon_i \tag{39}$$

Where $E(R_i)$ is the expected return on asset *i*. R_f is the risk-free rate of return, taken as the 3-month US treasury bill rate. β_i is the systematic risk of asset *i*, sometimes known as just the risk of asset *i*. $[E(R_m) - R_f]$ is the market price of risk, the excess return above the risk-free rate that holding the entire market portfolio is able to achieve. γ_i is the factor loading for SMB (Small Minus Big). SMB being the factor that notes the difference between the outperformance of small firms compared to large firms. δ_i is the factor loading for HML (High Minus Low). HML being the factor that notes the difference between the outperformance of growth stocks compared to value stocks. η_i is the factor loading for MOM (Momentum). MOM being the factor that notes the difference between the outperformance of stocks showing short term momentum over stocks without short term momentum. ε_i is the error term to account for inconsistencies in the model. The model, as in research question 2, is standard in adjusting for risk premium.

I again implement the equal weighting strategy, where the composition of the portfolios has an equal weighting of each stock. In terms of study objectives of short interest, the Fama and French Three Factor Model with Momentum shows us where in particular the returns are coming from based on the previously mentioned four factors of market risk premium, SMB, HML and MOM. The α_i will show us the under and overperformance of my model in relation to short interest, with a high α_i, we will see that my model is overperforming based on either high or low short interest portfolios while a low α_i will show that my model is underperforming based on either high or low short interest portfolios. Based on previous literature I expect the underperformance of heavily shorted stock portfolios (shown with a lower α_i than lightly shorted portfolios).

Regarding the regression I will again be checking for multivariate normality, no multicollinearity and homoscedasticity in the results. I will perform tests in line with research question 1 to check for these conditions and to amend the dataset if necessary to counteract these conditions.

As far as I am aware the Toronto Stock Exchange has not implemented any shorting bans between February 2010 and July 2017. If a ban were to exist those months would be excluded from my sample in line with research question 2, Boehmer et al. (2010) and Desai et al. (2002).

It is worth noting that since I have a smaller dataset, roughly of around 280 firms per month, the portfolios formed will be smaller than the portfolios formed in research question 2 and in particular the portfolios formed by Boehmer et al. (2010). This is natural since the Toronto stock exchange has fewer listings than the NYSE and Nasdaq, the NYSE and Nasdaq also being home to dual listings[65] of many global multinationals.

[65] Dual listings are often very important for securities that are difficult to obtain from foreign markets. A lot of markets are not accessible for retail investors due to the process in which securities are cleared and other counterparty risk, therefore a listing of a stock in the US markets means most retail investors can access the stock. This is good in terms of liquidity and thus market efficiency. A lot of stocks prefer to list on liquid markets such as the London Stock Exchange, New York Stock Exchange and NASDAQ. ADRs (American Depositary Receipts) and GDRs (Global Depositary Receipts) are a common way to gain a foreign listing.

Refer in addition to Appendix 4 for a methodology diagram outlining the estimation procedures implemented to answer research question 3. The methodology diagram allows us to see the step by step processes that need to take place in the method to obtain the results.

5.4 Whether and to what extent short sales affect liquidity, price discovery, volatility and cross-section of stock returns? (Research Question 4)

This question aims to explore the relationship between short selling and its effect on volatility, liquidity and price discovery. I first outline the method to measure volatility, second the method to measure liquidity and finally the method to measure price discovery. The best way to measure these metrics is over periods where short selling has been banned and compare it to periods where short selling has not been banned. In this case the 2008 short sale ban for UK financial stocks is used to explore this.

5.4.1 Measuring Volatility

Volatility is the change in stock prices on a day to day basis created by the constant demand of buyers and constant supply of sellers. Volatility is known for clustering, where periods of high volatility occurs for a period only to be followed by periods of low volatility. The clustering nature means GARCH models are often a good fit to model volatility and can be amended to take note of changes in volatility for 2 or more sets of defined periods. What I find in my data is that I see periods of volatility clustering, which is normal for a financial time series.

However as good as GARCH models may be, they fail to consider the asymmetric nature of news on returns, where downside moves are more aggressive than upside moves. This can be seen as market participants having a risk-adverse nature naturally over a risk-seeking nature. An EGARCH model can take this into account and will better model datasets where this asymmetric nature of news on returns is present. From observing my daily returns, I see this asymmetric nature does not occur in the Unbanned Portfolio so GARCH is a better fit for both portfolios overall.

Picking the choice of model is based on your data type and historical tendencies of the data in terms of movement, but GARCH models require fitting based on GARCH and ARCH lags.

Fitting a model to the data is a main element of GARCH models as both ARCH and GARCH terms can take an infinite number of lags. The most common way an GARCH model can be fit is using either AIC (Akaike Information Criterion) or BIC (Bayesian Information Criterion). A series of models is run on econometric software like STATA or EViews such as GARCH (1,1), GARCH (1,2), …., GARCH(p,q). From these models the AIC and BIC are evaluated to judge the fit of the model with respect to the dataset. GARCH (3,3) is usually taken as my last estimate, to make computation feasible, as Maximum Likelihood Estimation can be computationally challenging at larger lags. The lowest value of AIC or BIC is taken as the best fitting model. It is also worth noting since Maximum Likelihood Estimation is used, convergence is not guaranteed, so some models may not fit at all. I use both the AIC and BIC for both Banned and Unbanned Portfolios.

Table 10: Unbanned Portfolio and Banned Portfolio Model Selection Using AIC and BIC Criterion

	Unbanned Portfolio (Control)		Banned Portfolio	
	(1)	(2)	(3)	(4)
Model	AIC	BIC	AIC	BIC
GARCH (1,1)	-1656.672	-1635.217	-1224.232	-1202.776
GARCH (1,2)	-1655.701	-1630.669	-1222.580	-1197.549
GARCH (1,3)	-1655.260	-1626.652	-1223.767	-1195.160
GARCH (2,1)	-1654.907	-1629.876	-1223.026	-1197.994
GARCH (2,2)	-1653.684	-1625.077	-1226.663	-1198.056
GARCH (2,3)	-1658.476	-1626.293	-1225.048	-1192.865
GARCH (3,1)	-1654.597	-1625.989	-1221.052	-1192.445
GARCH (3,2)	-1652.766	-1620.583	-1217.962	-1185.778
GARCH (3,3)	-	-	-1224.456	-1188.697

Notes: Table 10 shows the AIC and BIC criterion of model selection from GARCH (1,1) up to and including GARCH (3,3) for both Banned and Unbanned Portfolios. Models are fit with AIC and BIC criterion using Maximum Likelihood Estimation in STATA 15. Blanks denote cases of where the AIC and BIC criterion were not able to generate a fit, due to a flat log likelihood. Lowest level of AIC or BIC is taken as the better fit.

Table 10 shows the results for both Banned and Unbanned Portfolios in regards to AIC and BIC. We can see that AIC and BIC are reasonably correlated with each other, which is not a surprise since both computations are based off the likelihood function. In regards of the Unbanned Portfolio the best fitting model is the GARCH (2,3) model with AIC and the GARCH (1,1) with BIC. With the Banned Portfolio, the best fitting model is the GARCH (2,2) model with AIC and the GARCH (1,1) model with BIC. This is a likely situation where two sets of data are going to produce better fits in certain models over another, in particular with different criterions such as AIC and BIC. To remain consistent in my study, I decide to use the GARCH (1,1) model for both portfolios. In terms of BIC criteria, this is the best fitting model for both sets of data. In terms of AIC criteria, it is the second-best fitting model for the Unbanned Portfolio and the fourth best fitting model for the Banned Portfolio. This in turn makes comparisons between both portfolios easier and it also means emphasis of fit is put on the Banned Portfolio and the Unbanned Portfolio. We are indeed investigating the effects of a short sale ban on volatility, but we must also be aware of the fit we achieve with respect to my control variable.

I model volatility using a GARCH (1,1) model. This is the simplest model in the GARCH family using 1 lag for the ARCH term and 1 lag for the GARCH term. More details of the construction of GARCH models can be found in section 3.1.4 under the theoretical framework chapter of this thesis. I use an GARCH (1,1) model to measure the volatility of 12 banned stocks and 12 unbanned stocks for the periods of the short sale ban and no short sale ban. The 12 banned stocks are my independent variables and the 12 unbanned stocks are my control variables, I employ the same GARCH (1,1) model for both sets of stocks. My study period is from 3rd January 2008 to 16th January 2009, the short sale ban occurs from 19th September 2008 to 16th January 2009.

I employ a GARCH (1,1) model to overcome heteroscedasticity that exists in the daily return data for my 24 stocks. If I did not have heteroscedasticity I could use the standard deviation of the abnormal return instead. The GARCH (1,1) model is built from the CAPM in my case, where the error term of the CAPM is the GARCH part of the model accounting for the volatility clustering.

$$R_t = \alpha_t + \beta_i R_{mt} + \varepsilon_t \tag{40}$$

$$\varepsilon_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + D_{SSB} \tag{41}$$

The model shown in equations (40) and (41) is the GARCH (1,1) model with a dummy variable used to model volatility in both Banned and Unbanned Portfolios. Equation (40) is the mean equation and equation (41) is the variance equation. R_t is the daily average return of unbanned or banned stock portfolios. R_{mt} is the daily return of the FTSE 100, taken as the market return. ε_{t-1} is the first ARCH value, this is the past squared residual which represents the effect of recent news to future stock volatility. σ_{t-1}^2 is the first GARCH value, which represents the lagged value of variance to capture the long-term effects of volatility. D_{SSB} is the dummy variable, which takes the value of 0 or 1 to represent the short sale ban and short sale no ban periods respectively.

The coefficient of the dummy variable is used to determine whether there is a change in volatility and what direction that change in volatility is. If the coefficient is negative (or positive) and there is statistical significance at the 5% level it can be seen that there is an increase (or decrease) in volatility respectively for the portfolio of banned or unbanned stocks. However, if the coefficient is 0, and holds statistical significance at the 5% level, we can say that there is no change in volatility. The size of the coefficient will determine the degree of change in volatility, this can be especially useful when comparing Banned and Unbanned Portfolios with each other.

5.4.2 Measuring Liquidity

Liquidity is the ability to buy and sell a security in ease in which to not affect the bid or ask price of a security too dramatically. Illiquid stocks have wider bid-ask spreads since market makers need a potential for higher profit in order to be influenced in holding a riskier security (in the sense it is harder to sell or buy). The bid-ask spread is often used as one indicator of liquidity in a security, the wider the bid-ask spread the more illiquid a security is seen to be and vice versa.

I use a bid-ask spread model as in Lobanova et al. (2010) to measure the liquidity in both Banned and Unbanned Portfolios for no short sale ban and short sale ban periods.

$$S_t = c + \beta_0 \alpha r_t^2 + \beta_1 \alpha v_{it} + \beta_2 e v_{it} + \beta_3 D_{SSB} + \varepsilon_t \tag{43}$$

The regression model shown in equation (43) is a Bid-Ask Spread Model that has been used in Lobanova et al. (2010) to measure liquidity. S_t is the average spread for Unbanned or Banned Portfolio at time *t*, αr_t^2 is the daily average return of the Unbanned or Banned Portfolio squared at time *t*, αv_{it} is the daily average volume for the Unbanned or Banned portfolio at time *t*, ev_{it} is the excess daily average volume of the Unbanned or Banned Portfolio at time *t*. The excess daily trading volume is calculated by subtracting the daily trading volume at time *t* from the average daily trading volume from the duration of January 2008 to January 2009. D_{SSB} is the dummy variable which takes the value of 0 or 1 for short sale ban or no short sale ban period being compared. If the coefficient of D_{SSB} has a negative value and is significant it can be proved there is a decline in liquidity from the first no short sale ban period to the second short sale ban period, if the coefficient of D_{SSB} has a positive value and is significant it can be proved there is an increase in liquidity from the first no short sale ban period to the second short sale ban period. A coefficient of 0 and significance for D_{SSB} indicates there is no change in liquidity. This is the basis of the model proposed by Lobanova et al. (2010).

This Bid-Ask Spread Model is a form of regression analysis using OLS and therefore needs to meet OLS conditions in order to remain BLUE. S_t is the independent variable in my regression, αr_t^2, αv_{it} , ev_{it} and D_{SSB} are the dependent variables being regressed on the independent variable. c is the intercept and ε_t is the error term. If the model is correctly specified ε_t will have a mean of zero.

Often it is difficult to estimate S_t in equation (44) given that bid and ask data can be difficult to obtain for securities in emerging markets and even for securities in developed markets. Therefore, it is best to use an estimator as a proxy for S_t . To address this problem, Corwin and Schultz (2012) propose a method to estimate the bid-ask spread using daily high and low prices for securities, which I do for my 24 stocks. The Corwin and Schultz (2012) bid ask spread estimator is as follows:

$$S = \frac{2(e^{\alpha}-1)}{(1+e^{\alpha})} \tag{44}$$

Where:

$$\alpha = \frac{\sqrt{2\beta}-\sqrt{\beta}}{3-2\sqrt{2}} - \sqrt{\frac{\gamma}{3-2\sqrt{2}}} \tag{45}$$

$$\beta = E\left\{\sum_{j=0}^{1}(ln\frac{H_{t+j}^{0}}{L_{t+j}^{0}})^2\right\} \tag{46}$$

$$\gamma = E\left\{\sum_{j=0}^{1}(ln\frac{H_{t,t+1}^{0}}{L_{t,t+1}^{0}})^2\right\} \tag{47}$$

By having this spread estimator, I can use high and low prices to work out a good approximation to spreads for my 24 stocks. High and low prices are readily available for these 24 stocks. From equations (44), (45), (46) and (47), α is the high-low price estimator, β is the summation of the squared high-low price spread. γ is the squared high-low price spread. S is the stock's high-low bid-ask spread estimator. H_t is stock's high price on day t, L_t is the stock's low price on day t. $H_{t,t+1}$ is the stock's highest price over the two-day period of t and t+1. $L_{t,t+1}$ is the stock's lowest price over the two-day period of t and t+1.Corwin and Schultz (2012) bid ask spread estimator is good in the sense we only require high and low prices for a security over a two-day rolling period, which is much easier to obtain than actual bid-ask prices. Corwin and Schultz (2012) spread estimator has been seen to be 10% within the true spread according to Corwin and Schultz (2012).

5.4.3 Measuring Price Discovery

Price discovery is the ability of a market to determine the fair price of a security efficiently and without delay. If price discovery is hampered, over and under-pricing of securities may occur for longer than in efficient markets. This is an advantage to speculators in the market, as they can take advantage of this if they can identify it. Generally, markets have better price discovery when less restrictions are imposed as seen in Brockman and Hao (2011) and Sochi and Swidler (2018), so that supply and demand is allowed to fully incorporate all known information. By restricting the natural supply of stocks, natural demand of stocks and the availability of information on stocks, price discovery can be impacted negatively. A deterioration in price discovery hampers market efficiency. Less efficient markets overall hurt buyers and sellers and can decline confidence in a market. Market confidence declines can lead to general economic declines, as negative sentiment can lead firms to slow down investment and consumers to stop spending. This in turn hurts economic growth. Overall, we can see having good price discovery in a market stretches beyond market participants, it can affect the general economy as a whole through a chain of effects.

For price discovery I can use a non-parametric test known as a Wald Wolfowitz Runs Test (or more commonly known just as a Runs Test) to explore randomness in stock returns. This test was proposed by Abraham Wald and Jacob Wolfowitz, both prominent statisticians and has been applied across many studies such as Molik and Bepari (2009) and Sochi and Swidler (2018) to measure price discovery. For this test I use the daily returns of the Banned and Unbanned Portfolios. A run is defined as a string of daily returns that are of the same sign. Zero also continues the run, and the run only ends where there is a sign change, either positive or negative.

I can show how a run is calculated using this example. We can use *P* to denote a daily positive return in a portfolio, *N* to denote a daily negative return in a portfolio and 0 to show no change in daily return of a portfolio. Using this information, we can develop the following string of runs as an example.

P N PP0PP NN PPP N P

We can see that we have a string of 14 daily portfolio returns, which includes 7 separate runs (spaces have been used to show the start of a new run and the end of an old run). Knowing this we can see that the shortest runs are a 1-day negative run occurring twice and a 1-day positive run occurring twice. The longest run is a 5-day positive run occurring once.

A related test is the Kolmogorov-Smirnov Test and can also be applied instead of the Wald-Wolfowitz Runs Test. The Kolmogorov-Smirnov Test is more powerful when looking at the difference between distributions looking at their locations, however the opposite case holds when the distributions differ in variance and have only a small difference in location.

The samples being explored are the pre-ban period for the Banned Portfolio, the ban period for the Banned Portfolio, the pre-ban period for the Unbanned Portfolio and the ban period for the Unbanned Portfolio.

To understand the link between the run tests and short selling effect on price discovery we have to look at the tails of the run distribution. If the tails of the run distribution are larger than expected by the unbanned stock portfolio, it would indicate the short selling ban impacting price discovery. If a short selling ban inhibits price discovery given bad news, then the returns would be more negatively skewed than we would find in an efficient market. This means more runs would be found in the left tail of the distribution. Also, if good news causes markets to overreact, a short selling ban might stop a correction, this means the right tail of the distribution would have more runs. Therefore, fatter left and right tails of the run distribution are indicative of a short sale ban affecting price discovery.

It is very much evident when exploring a short-sale ban period with a control portfolio (the Unbanned Portfolio), that these changes in runs are caused by the short sale ban and no other outside factors. Islam and Khaled (2005) note that emerging markets in particular are characterised by lower volumes, ability of manipulation of stock prices by a few large traders, less stringent accounting requirements, delays in settlement and the weaker transmission of public financial information. Thus, in emerging markets, price discovery issues may also arise from these problems, therefore exploring a market with a control portfolio is essential to mitigate outside factors. I in turn do this in a developed market of the United Kingdom, thus making my results less prone to outside factors affecting price discovery even more.

In my method, I look at the distribution of *n* day runs by dividing the 4 samples into runs of 5 days or longer and runs between 1 and 4 days. I use a percentile on the distribution to show why 5 length runs are considered long and why 4 length runs are considered normal. Achieving more runs than expected that are 5 days or longer indicates fat tails, and implies a short selling ban restricts price discovery. This means there are extended periods for markets to understand new information. If that result is obtained, it necessarily follows that shorter run periods (1–4 days) will have fewer runs than expected in an efficient market. In my case the Unbanned Portfolio is used as a control for market efficiency. A comparison between the independent variable (Banned Portfolio) and control variable (Unbanned Portfolio) is sufficient to show the effect of the short sale ban on price discovery.

These three methods allow me to measure volatility, liquidity and price discovery between the two time periods (short sale ban or no short sale ban) in question. I repeat both methods for my banned stocks and unbanned stocks, which is essential as I am employing a control. This means we can be confident that it is in fact the short sale ban that is affecting liquidity, volatility and price discovery and not an outside third factor.

Refer in addition to Appendix 5 for a methodology diagram outlining the estimation procedures implemented to answer research question 4. The methodology diagram allows us to see the step by step processes that need to take place in the method to obtain the results.

5.5 Conclusion

For research questions 1 to 3 the estimation methods are similar and are solid on an econometric basis. I form lightly shorted and heavily shorted portfolios and regress excess returns on asset pricing models. This technique has been observed in the past across many empirical studies in the short interest literature such as Boehmer et al. (2010), Asquith et al. (2005) and Desai et al. (2002). I believe given this I am robust in my techniques.

On an econometric basis I am conducting a multiple linear regression with the excess return of a portfolio being the independent variable and the sensitivity to my asset pricing factors being my dependent variables. With multiple linear regression we need to hold a certain amount of assumptions for my results to be significant. Such assumptions include that of multivariate normality, no multicollinearity and homoscedasticity, I account for these assumptions also in my estimation methods. I perform the respective tests to take into consideration these assumptions and ensuring my models are BLUE. Given I take a robust method and apply a sound econometric framework, I can have confidence in my methods.

My estimation methods for research questions 1 to 3 involve a great degree of complexity and are time intensive. I am dealing with over 180,000 unique observations in some cases, there is a dynamic nature to my portfolio forming technique since month by month the portfolios will change composition. Stocks also exit and enter the exchanges as a whole (in terms of IPOs and delisting) and this means it is very rare that two portfolios each month will be alike. The construction of the portfolios is a manual task and portfolios have to be constructed each month manually, I cannot employ a formula or repetitive method to construct portfolios since the number of stocks in the market as whole change on a month by month basis. Given this, I am still adamant that by using such a technique I can gain results of great value. I am aware that the construction of the datasets in the single biggest challenge of research questions 1 to 3 over the actual estimation methods. The ability to run a regression can take seconds, however the manual construction of the portfolios can take weeks.

Regarding question 4 I employ new methods not seen in questions 1 to 3, I am concerned with comparing changes in liquidity, volatility and price discovery over two distinct periods of time using two stock portfolios (unbanned and banned). I am investigating the effect of the short sale ban in 2008 on stock liquidity, volatility and price discovery. I use a Banned Portfolio in order to investigate this effect and use an Unbanned Portfolio as a control.

The methods I employ for question 4 involve using a GARCH (1,1) with a dummy variable to see changes in volatility, GARCH models are ideal because of the volatility clustering phenomenon that occurs in finance. To fit the GARCH model, I employ the AIC and BIC criterions, which allow us to see which models fit the data better, so in turn we are not fitting a model that is too distinct from the data. It is natural that no model will fit perfectly, but it is essential we minimise as much error as we can. To measure changes in liquidity I use a Bid-Ask Spread model from Lobanova et al. (2010) that again uses a dummy variable to see changes between two time periods. I employ a Bid-Ask Spread estimator using high and low stock prices due to the lack of availability in bid-ask prices for all securities from the London Stock Exchange. The Bid-Ask Spread model is a form of regression analysis using the OLS estimator and therefore needs to meet the conditions of OLS in order to be BLUE. To measure change in price discovery, I use a Wald-Wolfowitz Runs Test to look for fat tails in the run's distribution. Fat tails are indicative of market efficiency being impacted due to a short sale ban.

CHAPTER 6: EMPIRICAL RESULTS

The results chapter sums up my findings and provides analysis to my research questions outlined at the start of this thesis. At the start of my thesis we began with four research questions. I will again list my research questions for reference at start of each respective section in Chapter 6. My results will aim to directly answer these questions in particular, while giving a broader outlook of my findings.

My results have again been divided up into sections depending on the research question, this is to keep the results, methodology and literature review sections separate rather than taking a multiple paper approach. Many of these research questions interlink, such as research questions 1 to 3 and this adds to the flow of the study.

In particular, results in section 6.1 report for research question 1, results in section 6.2 report for research question 2, results in section 6.3 report for research question 3 and results in section 6.4 report for research question 4. Subsections are also included to break up results for research questions investigating different lines of enquiry using the same dataset.

Chapter 7 lists a full conclusion based on my findings in Chapter 6.

6.1 Five Factor Model as an Adjuster for Risk Premium (Research Question 1)

This research question aimed to answer the following question: In adjusting for risk, Boehmer et al. (2010) uses the Fama and French (1993) three-factor model augmented by the momentum factor. However, is this long/short strategy still valid if a different and more recent model, such as Fama and French (2015) five-factor model, is used to adjust for risk premium?

I first report the descriptive statistics of the individual portfolios formed before the regression for these portfolios is run in **Table 11**, the main results for the regression are reported in **Table 12**.

Table 11: Descriptive Statistics for Portfolios

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Portfolios	Mean	Standard Error	Median	Standard Deviation	Kurtosis	Skewness	Range	Maximum	Minimum
SIR 95%	0.019	0.008	0.016	0.085	3.356	0.338	0.628	0.393	-0.235
SIR 90%	0.015	0.008	0.011	0.083	3.225	0.258	0.624	0.373	-0.251
SIR 5%	0.031	0.006	0.032	0.060	2.068	-0.449	0.396	0.213	-0.183
SIR 10%	0.026	0.005	0.028	0.054	2.671	-0.719	0.361	0.168	-0.193
SIR 5%-SIR95%	0.012	0.005	0.017	0.055	4.390	-0.441	0.438	0.204	-0.234
SIR 10%-SIR90%	0.011	0.005	0.016	0.047	3.631	-0.823	0.347	0.142	-0.205

Notes: Data source using Microsoft Excel from Thomson Reuters US Total Return Index DataStream Data. Table 11 shows the descriptive statistics for the individual portfolios formed before regression including the standard and long/short portfolios. I report the mean, standard error, median, standard deviation, sample variance, kurtosis, skewness, range, minimum and maximum values of the portfolios. The descriptive statistics above given the distribution of the return of the portfolios on a total return basis. Data author's own calculation.

Table 11 shows the descriptive statistics for portfolios run in the regression. The mean shown in Column (1) will be the same as the raw return and represents the average return each portfolio produces on a monthly basis. We can see that the highest mean lies with the 5th percentile portfolio and this produces the best return on a monthly basis. The standard error shown in Column (2) tells us the deviation of the sample mean from the population mean, as in the confidence of my mean result. We are within very reasonable standard error rates for all portfolios, so we can be confident with my mean.

The median shown in Column (3) tells us the middle value of return in my portfolios, this is a good indicator to look at when samples are skewed as heavy skew can distort the mean as an indicator of average. The median in relation to the mean can tell us the skew of the data. We can see that the return of the heavily shorted portfolios is skewed to the left, while the return of the lightly shorted portfolios is skewed to the right. The median being the same as the mean would indicate little to no skew in the data.

The standard deviation shown in Column (4) tells us the deviation of the mean, where we can see the heavily shorted portfolios have the greater standard deviation than the lightly shorted portfolios.

Kurtosis shown in Column (5) is the measure of the distribution of the tail of a sample, how many more extreme observations are contributing to the sample as opposed to non-extreme observations. The Kurtosis is much higher with the heavily shorted portfolios as compared to the lightly shorted portfolios, there are many portfolio returns in the extremes especially during periods of crisis in the market such as the 2008 financial crash.

The Skewness shown in Column (6) is the measure of the distribution of returns in a sample from the mean, whether returns are mainly towards the right or left of the mean. We exhibit positive skew for my heavily shorted portfolios and negative skew for my lightly shorted portfolios. A potential reason for the opposite skewness in the heavily shorted portfolios compared to the lightly shorted portfolios shows us the monthly returns in lightly shorted portfolios are being generated by few months of extreme positive gains (skewing the monthly returns), while the monthly losses in heavily shorted portfolios are being generated by a few months of extreme negative losses (skewing the monthly returns). Therefore, we can say that the distribution of the monthly returns is skewed for both portfolios. The distribution of the median above or below the mean is also an indication of skewness and is also indicated in Table 11. It is very rare to have perfectly symmetrical data, so an element of skewness will usually always be present. These skewness results are based on my findings and consistent with the existing literature such as Desai et al. (2002) and Boehmer et al. (2010).

The Range shown in Column (7) shows the difference between the maximum and minimum return per month. The Maximum shown in Column (8) shows the maximum return in a portfolio in a month. The Minimum shown in Column (9) shows the minimum return in a portfolio in a month. We again see a greater range in the heavily shorted portfolios as opposed to the lightly shorted portfolios. This is also reflected in the lower minimum levels and higher maximum levels for the heavily shorted portfolios.

In general, we see that my lightly shorted portfolios are much more stable when it comes to predictability of returns. The volatility lies with the heavily shorted portfolios which can have excessive up and down moves on a month by month basis based on risk aversion and risk taking. The heavily shorted portfolios have a large degree of short interest in them and are susceptible to short squeezes, thus amplifying moves on the upside. It is advisable to hold lightly shorted portfolios to have a lower volatility profile in an investor's overall portfolio.

Table 12: Regression Analysis Results of Monthly Equal Weighted Returns on Highly and Lightly Shorted Stock Portfolios for the United States

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Portfolios/#Stocks	Raw Return	Excess Return	Intercept	EM-RF	SMB	HML	RMW	CMA
SIR 95%	0.019	0.017	0.013	0.128	0.313	0.641*	-0.176	0.159
#64 stocks			(0.124)	(0.602)	(0.314)	(0.056)	(0.675)	(0.719)
SIR 90%	0.015	0.014	0.010	0.112	0.312	0.683**	-0.154	0.004
#127 stocks			(0.233)	(0.639)	(0.304)	(0.038)	(0.708)	(0.993)
SIR 5%	0.031	0.029	0.025***	0.236	0.317	0.142	0.082	0.461
#64 stocks			(0.000)	(0.174)	(0.151)	(0.548)	(0.782)	(0.145)
SIR 10%	0.026	0.024	0.019***	0.209	0.376**	0.250	0.188	0.299
# 127 stocks			(0.000)	(0.173)	(0.055)	(0.233)	(0.474)	(0.284)
SIR 5%-SIR 95%	0.012	0.012	0.012**	0.108	0.004	0.500**	0.258	0.301
			(0.037)	(0.506)	(0.986)	(0.026)	(0.358)	(0.311)
SIR 10%-SIR 90%	0.011	0.010	0.009**	0.097	0.063	0.433**	0.342	0.295
			(0.068)	(0.487)	(0.722)	(0.025)	(0.158)	(0.249)

Notes: Table 12 shows the regression analysis of monthly equal weighted highly and lightly shorted stock portfolios. Source from Microsoft Excel from Thomson Reuters US Total Return Index DataStream Data. The regression equation run to obtain each portfolio in the table is as follows:

$$R_{port\,t} - R_{ft} = \alpha_{port} + \beta_{port}\left(R_m - R_f\right)_t + \gamma_{port}SMB_t + \delta_{port}HML_t + \eta_{port}RMW_t + \lambda_{port}CMA_t + \varepsilon_t \qquad (48)$$

The dependent variable is the equal weighted monthly portfolio excess return $R_{port\,t} - R_{ft}$ on highly or lightly shorted stock portfolios for the month subsequent to portfolio creation on stocks from the Thomson Reuters US Total Return Index. $\left(R_m - R_f\right)_t$ is the market risk premium, SMB_t is the excess return of a portfolio of small cap stocks over a portfolio of large cap stocks, HML_t is the excess return of a portfolio of high book to market stocks over low book to market stocks, RMW_t is the excess return of a portfolio of robust profitability stocks over weak profitability stocks and CMA_t is the excess return of a portfolio of conservative stocks over aggressive stocks. The monthly data for the five Fama French factors is from Kenneth French's Data Library. The portfolios SIR 95% and SIR 90% include stocks with short interest ratios from the 95th and 90th percentile respectively in month t-1. The portfolios SIR 5% and SIR 10% include stocks with short interest ratios from the 5th and 10th percentile respectively in month t-1. The portfolios are rebalanced each month based on new short interest data. Portfolio raw returns and excess returns over the risk-free rate are shown under the headings of Raw Return and Excess Return. The number of stocks in each portfolio can change each month and thus the average number of stocks in each portfolio is shown under the heading # stocks. I report p values in brackets under respective regression coefficients. Stars are used to indicate significant figures. * is significance at the 10% level, ** is significance at the 5% level and *** is significance at the 1% level. My sample period is based on monthly short interest data from January 2001 to January 2010. Data author's own calculation.

I next report the regression analysis results in **Table 12** based on equation (48). If we look at Column (1) and (2) of Table 12 we observe that the best performing standard portfolio is the SIR 5% portfolio with a raw return of 3.1% per month and an excess return of 2.9% per month above the risk-free rate. The worst performing standard portfolio is the SIR 90% portfolio with a raw return of 1.5% and an excess return of 1.4% per month above the risk-free rate.

Both SIR 5% and SIR 10% portfolios outperform the SIR 95% and SIR 90% portfolios on both a raw and excess return basis. Lightly shorted stocks do indeed outperform heavily shorted stocks according to my results. The outperformance of lightly shorted stocks over heavily shorted stocks is consistent with Boehmer et al. (2010), Desai et al. (2002) and Asquith et al. (2005). This outperformance has been noted by the previous authors due to the fact that the market is a good predictor of future stock returns and the increase in short interest reflects that. It must be noted that the distribution of outperformance is not even since the SIR 90% portfolio has a lower raw and excess return than the SIR 95% portfolio.

There is a good difference between raw and excess returns in standard portfolios, since before the 2007-2009 financial crisis interest rates from central banks were very much normal on a historical basis[66]. Given the excess return is the difference between the portfolio return and the risk-free rate, it is natural for there to be a good difference if the risk-free rate is much higher than the general lower bound of zero.

Moving to Column (3) of Table 12, we look at the intercepts of the regression for the standard portfolios. The intercepts indicate what the Fama and French Five Factor Model has failed to specify, or more commonly known as the alpha of the model. A positive intercept coefficient will indicate that the return data has performed better than the model expected and a negative intercept coefficient will indicate that the return data has performed worse than the model expected. All four standard portfolios hold a positive intercept coefficient.

The intercepts of the SIR 5% and SIR 10% portfolios (0.019 and 0.025 respectively) are much higher than the intercepts of the SIR 95% and SIR 90% portfolios (0 for both portfolios based on statistical significance). This indicates that lightly shorted stocks outperform heavily shorted stocks on a risk adjusted alpha basis as well for my time period.

Moving onto the coefficients of my regression, Column (6) shows that in the standard portfolios, the heavily shorted portfolios exhibit a higher positive coefficient on HML compared to their lightly shorted counterparts. Column (7) shows that both heavily shorted standard portfolios and lightly shorted standard portfolios have a zero coefficient on RMW based on statistical significance. Moving to Column (8), the CMA coefficient is zero on my standard lightly shorted portfolios and is zero on my standard heavily shorted portfolios based on statistical significance. The SMB factor in my standard portfolios is statistically significant only in the case of the SIR 10% portfolio, this is in contrast to Boehmer et al. (2010), although my periods of study are not identical so this may be one of the reasons for this difference.

Overall both the estimated coefficients of RMW and CMA factors do not hold statistical significance, showing that in my case those factors are not influencing the independent variable of the model. This brings into doubt the effectiveness of those two factors and whether regressing without them would lead to a better fit of model. The HML difference in the standard portfolios indicates that lightly shorted stocks often exhibit qualities of value stocks and heavily shorted stocks often exhibit qualities of growth stocks. The distinction between growth and value stocks being based on the price to earnings multiple often just referred to as "the

66 Across my study the fed funds rate between 2001 and 2010 varied dramatically between years. In 2001 the fed funds rate was at 6.5%, which it was cut down to 1.75% in 2002 for the dot com bear market. Further cuts were made in 2003 to 1.25% and in 2004 to 1%. Fed tightening occurred from 2003 to 2006, which saw the fed funds rate rise to 5.25%. Between 2007 to 2009 the fed funds rate was cut from 5.25% to a target range of 0%-0.25%, an all-time low in fed history. Till 2010 the target range of 0%-0.25% was maintained. The fed funds rate is very much a proxy to the risk-free rate available to investors, as it affects short term treasuries such as the 3-month US treasury bond. The fed funds rate also affects the entire yield curve and demand for bonds and stocks.

multiple". The SMB factor is higher on my lightly shorted portfolio of SIR 10% and not statistically significant for other portfolios, this means my SIR 10% portfolio exhibits characteristics of the small firm effect, where small firms outperform big firms in the long term.

Moving onto the long/short portfolios, Column (2) shows that the monthly abnormal return of the SIR 5%-SIR 95% is at 1.2% per month. The portfolio also holds a zero-market beta, which is not of significance since a negative market beta is what Boehmer et al. (2010) would have been looking for. This means that if this portfolio with a negative market beta was to be added to an existing holding, returns would be boosted while at the same time market risk being reduced. Boehmer et al. (2010) proposed this strategy of going long the least shorted stock portfolio and short the most shorted stock portfolio, this can be modelled with the SIR 5%-SIR 95% portfolio. This was proposed because of the negative beta and the large abnormal return, we however do not propose this. It is worth noting that there is a fee often associated with shorting stocks, therefore making a direct comparison between long only and long/short portfolios is not as direct. Even if we include a shorting fee for the borrowed stocks, we find that the long/short portfolios underperform even more than what my results show. This in particular is highlighted in Lu et al. (2018) where the cost of borrowing can sometimes erase substantial profits of a long/short portfolio. In the most extreme of cases shorting costs can exceed 50% of the value of stock.

In my findings, the best performing portfolio by far has been the SIR 5% portfolio and it seems most viable to hold the SIR 5% portfolio and not apply a short strategy on the other end.

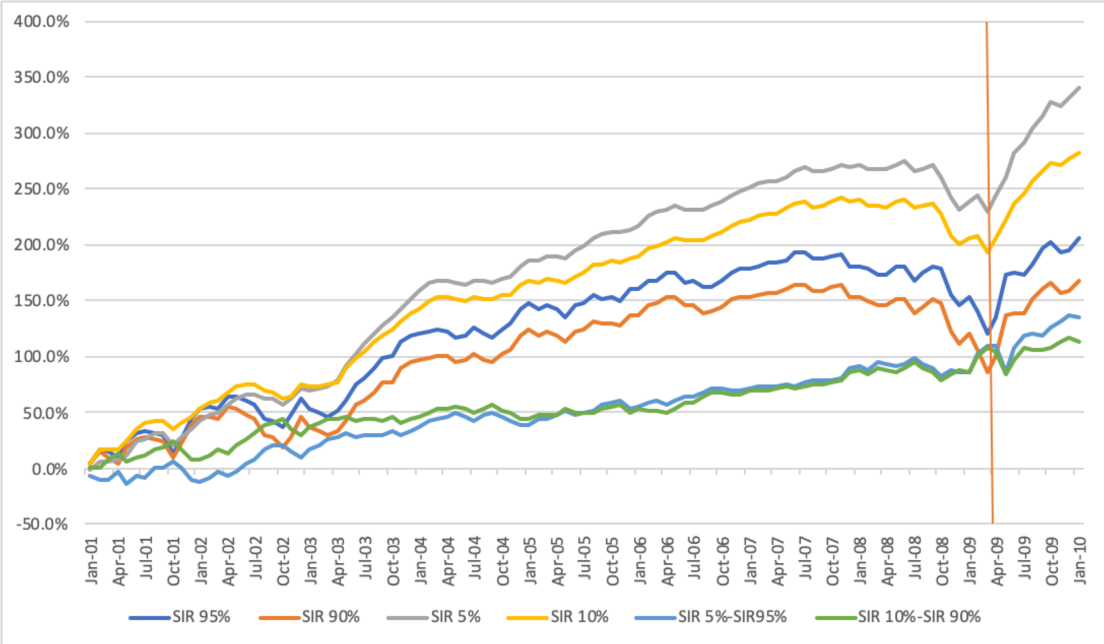
Table 13: Regression Statistics for Portfolios

	(1)	(2)	(3)	(4)
Portfolios	Multiple R	R Square	Adjusted R Square	Observations
SIR 95%	0.321	0.103	0.060	109
SIR 90%	0.319	0.102	0.058	109
SIR 5%	0.327	0.107	0.063	109
SIR 10%	0.373	0.139	0.097	109
SIR 5% - SIR 95%	0.225	0.051	0.005	109
SIR 10% - SIR 90%	0.234	0.055	0.009	109

Notes: Table 13 shows the regression statistics for each of the portfolios, with metrics including Multiple R, R Square, Adjusted R Square and Observations. Data author's own calculation. Multiple R shows the relationship between the independent variable and the multiple regressors for that independent variable. R Square is the square of the Multiple R and is often used as the main relationship indicator between the independent variable and regressor(s). R Square removes the negative sign when squaring. R Square is quoted between 0 and 1, where 0 is no relationship between independent variable and regressors and 1 is a perfect linear relationship between independent variable and regressors, it is very rare to see a R Square of 1, where higher R Squares are preferential in the hard sciences with lower R Squares acceptable in the social sciences. The Adjusted R Square is a modified version of R Square that has been adjusted for the number of predictors in the model. The Adjusted R Square increases only if the new term improves the model more than would be expected by chance and it decreases when a predictor improves the model by less than expected by chance. Observations is the number of observations in the dataset and therefore is my sample size. I take observations from January 2001 to January 2010 inclusive and this constitutes 109 months.

Table 13 shows the regression statistics for portfolios. Looking at Column (2) of Table 13 we see that the best performing portfolio in terms of R Square is the SIR 10% portfolio and the worst performing portfolio in terms of R Square is the SIR 5%-SIR 95% portfolio. We see a relatively similar R Square across portfolios although the lightly shorted portfolios hold the higher R Square compared to the heavily shorted portfolios. I find the R square to be low in this model and this could very well be contributed to the lack of fit of the RMW and CMA factors of the Fama and French Five Factor Model.

Figure 11: Cumulative SIR Portfolio Returns for the United States



Notes: Data source using Microsoft Excel from Thomson Reuters US Total Return Index DataStream Data. Figure 11 shows the cumulative returns of the different short interest ratio (SIR) portfolios over the period of the study. Red Vertical Line in Figure 11 shows the downtrend in cumulative SIR portfolio returns due to the 2007-2009 financial crisis.

Figure 11 shows the Cumulative SIR Portfolio Returns for the United States. Looking at Figure 11 we can see that there is a remarkable deal of correlation between the standard portfolios, in particular two dips in July 2002 and January 2009 stand out. We also see a great deal of correlation between the long/short portfolios. We can easily see that overtime the outperformance of the SIR 5% portfolio is evident and we can see that the SIR 5% portfolio outperforms the SIR 95% portfolio by a margin over 140%. The SIR 5% performance drastically outperforms the others, including the SIR 10% portfolio which has stocks that are included in the SIR 5% portfolio. The worst performing portfolio is the SIR 10%-SIR 90% portfolio. Both long/short portfolios underperform the other standard portfolios, this is due to the high short interest portfolios of SIR 95% and SIR 90% having a positive monthly raw return.

The 2007-2009 financial crisis is very evident from Figure 11 (as shown with a red vertical line in Figure 11), where we can see that all my standard portfolios posted a negative total return. The long/short portfolios fared better as they contain a short component that benefited from a market downturn. In general, there was no safe place being long in the equity market during that time period, other than a select few ultradefensive low beta stocks.[67] It is also worth noting that the recovery was very sharp and if an investor had ridden out the volatility the investor would have come out better on the other side. It has been documented by the founder of vanguard, Jack Bogle that during a strong bear market the best course of action has often been to do nothing. Excitement and expenses are seen an as investor's enemies and if an investor maintains a long-term investment horizon they will in the majority of cases do well. Maintaining a short position during the crisis would of course be better, but market timing can be especially difficult for professional money managers, let alone retail investors.

Table 14: Correlation Matrix for SIR Portfolio Returns for the United States

Portfolios	SIR 95%	SIR 90%	SIR 5%	SIR 10%	SIR 5%-SIR 95%	SIR 10%-SIR 90%
SIR 95%	1	0.989	0.762	0.821	-0.709	-0.792
SIR 90%	0.989	1	0.776	0.842	-0.676	-0.787
SIR 5%	0.762	0.776	1	0.960	-0.084	-0.261
SIR 10%	0.821	0.842	0.960	1	-0.218	-0.331
SIR 5%-SIR 95%	-0.709	-0.676	-0.084	-0.218	1	0.935
SIR 10%-SIR 90%	-0.792	-0.787	-0.261	-0.331	0.935	1

Notes: Table 14 shows the Correlation Matrix for SIR Portfolio Returns for the United States. This correlation matrix matches the return correlations between high, low and long/short SIR portfolios. A correlation of 1 indicates perfect positive correlation, a correlation of -1 indicates perfect negative correlation and a correlation close to 0 indicates little to no correlation.

Table 14 shows the correlation matrix for SIR portfolio returns for the United States. The correlation matrix allows us to see how correlated returns are between different portfolios. We see the strongest positive correlation between the SIR 90% and SIR 95% portfolios of 0.989; the SIR 5% and SIR 10% portfolios also hold a strong positive correlation of 0.960. This is to be expected since the SIR 10% portfolio includes stocks from the SIR 5% portfolio and the SIR 90% portfolio includes stocks from the SIR 95% portfolio. Even with this knowledge, the correlations remain positively strong and show returns may very well be influenced from stocks in the extreme percentiles.

The strongest negative correlation is with the SIR 95% portfolio and SIR 10%-SIR 90% portfolio of -0.792. I believe the short component contributes a great deal to this negative correlation, as the SIR 95% portfolio and SIR 10% have a fair deal of correlation in returns over the long term.

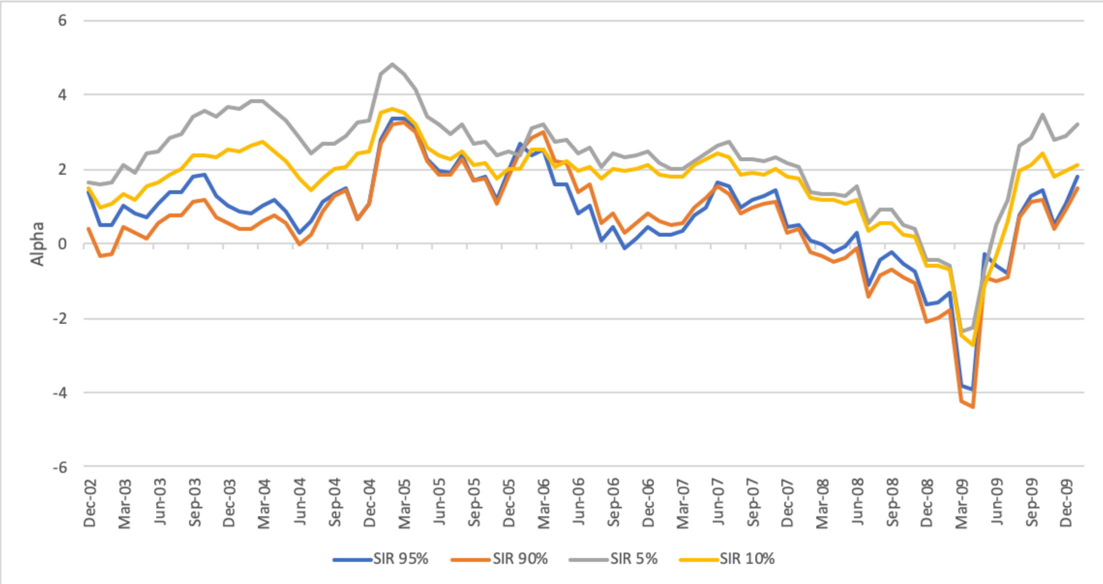
Equity markets experienced a great variety of changes over my sample period. Thus, in extension of my main results, I investigate if my findings are period specific or whether there are significant changes over time. To do this I conduct a 24-month rolling regression using the Fama and French Five Factor Model shown in Equation (48) to report for both alpha coefficient (abnormal returns) and beta coefficient (volatility with respect to market) across my sample period. I conduct the rolling regression for SIR 95%, SIR 90%, SIR 5%,

[67] A low beta stock would be one that has exhibited a beta close to zero. This means it is generally much less volatile than the market as a whole. These stocks are often termed defensive and usually perform best in the equity markets during a downturn. Of course, the best performance in a downturn would be found in fixed income safe haven assets such as government bonds, but in terms of equities low beta equities will outperform high beta equities usually in a recession.

SIR 10%, SIR 5%-SIR 95% and SIR 10%-SIR 90% portfolios, this means I account for heavily shorted portfolios, lightly shorted portfolios and long/short portfolios.

A rolling regression is a regression that shifts in time, in my case I will be moving ahead 1 month for the start and end date of the regression till the entire study period is covered. The length of 24 months is chosen as it provides a small time period to see specific differences. A rolling regression can be implemented with any model as long as the time period is in a discrete time interval.

Figure 12: SIR Portfolio Rolling Regression Alphas for the United States



Notes: Figure 12 shows the SIR Portfolio Rolling Regression Alphas for the United States. The rolling regression is conducted across a rolling 24-month period from January 2001 to January 2010 using the Fama and French Five Factor Model. The dates shown in Figure 12 account for the end period of that particular rolling regression, which is conducted on a monthly basis. I report for Alphas, which show the outperformance or underperformance of the portfolios with respect to the Fama and French Five Factor Model benchmark.

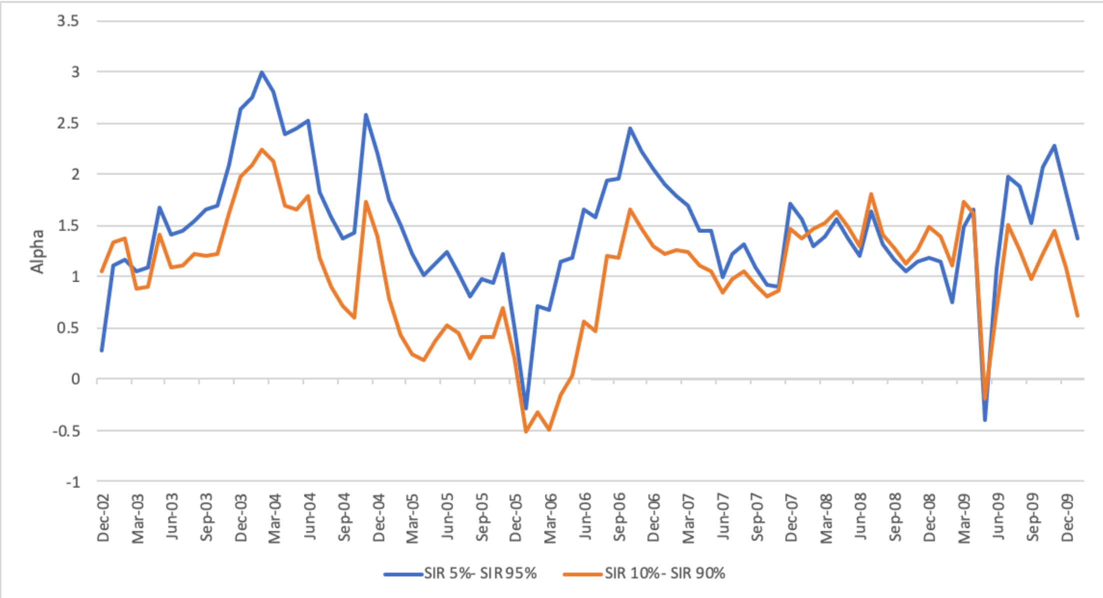
Figure 12 shows the rolling regression alphas for the United States using the SIR 95%, SIR 90%, SIR 10% and SIR 5% portfolios. We can see that in general the heavily shorted portfolios have lower alphas than the lightly shorted portfolios over all 24 month rolling regressions. There are few times the alphas overlap that of March 2005 and March 2006, but in general there is a clear distinction between the heavily shorted portfolio alphas and the lightly shorted portfolio alphas.

We see a great drop in alphas for both heavily and lightly shorted portfolios during the 2007 to 2009 financial crisis, which is evident since the mass short selling and panic selling contributed to much lower returns than any normal benchmark could identify. The alphas bottom in March 2009, corresponding to the bottom of the 2007 to 2009 bear market, the heavily shorted portfolios are much more affected than the lightly shorted portfolios in this case. This is backed up with studies such as Anderson et al. (2018) which show the underperformance of hedge fund alphas during recessions. The alphas do recover dramatically as the market recovers, showing that the 2007-2009 financial crisis was a short period of negative returns, that would have rebounded quickly if temperament of the individual investor was kept in check.

There is a good deal of correlation between the alphas of the heavily shorted portfolios and lightly shorted portfolios, showing a general risk-on and risk-off appetite in the market with respect to equities and other asset classes. When the market is going risk-off selling of equities is taking place across the board regardless of levels of short interest, when the market is going risk-on buying of equities is taking place regardless of levels of short interest. The market makes the distinction between heavily and lightly shorted portfolios, but I believe it also makes a distinction between equites and other fixed income asset classes.

The total range of alphas for the heavily shorted portfolios is between 3 and -4.2 and the total range of alphas for the lightly shorted portfolio is between 4.9 and -2.8. Range wise both heavily and lightly shorted portfolios are similar, but that lightly shorted portfolios naturally have a higher median and mean alpha as opposed to heavily shorted portfolios.

Figure 13: Long/Short SIR Portfolio Rolling Regression Alphas for the United States

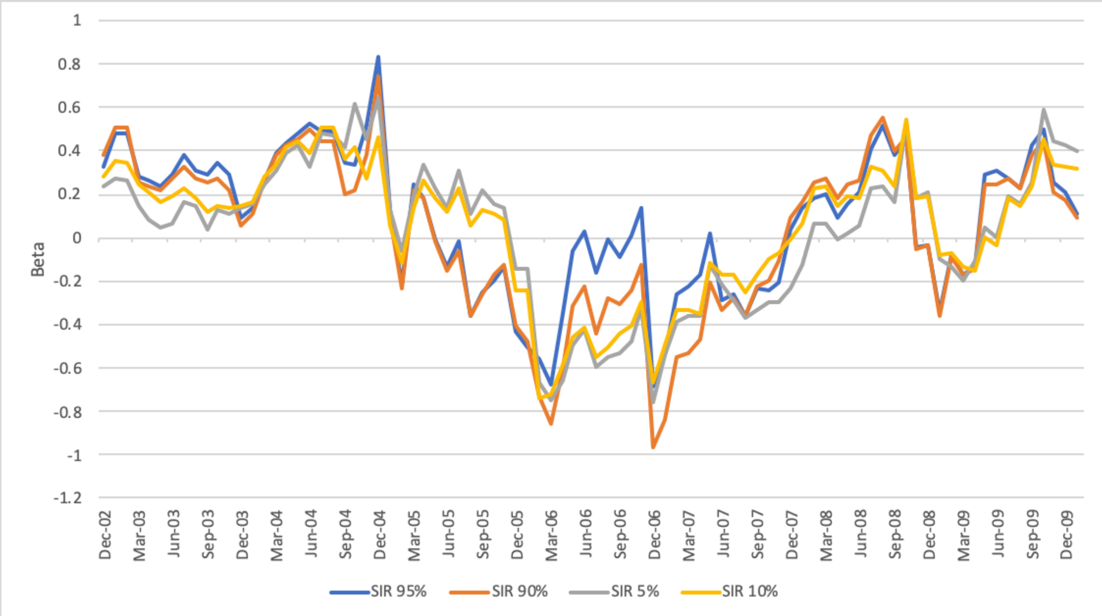


Notes: Figure 13 shows the Long/Short SIR Portfolio Rolling Regression Alphas for the United States. The rolling regression is conducted across a rolling 24-month period from January 2001 to January 2010 using the Fama and French Five Factor Model. The dates shown in Figure 13 account for the end period of that particular rolling regression, which is conducted on a monthly basis. I report for Alphas, which show the outperformance or underperformance of the portfolios with respect to the Fama and French Five Factor Model benchmark.

Figure 13 shows the rolling regression alphas for the United States using the SIR 5%-SIR 95% and SIR 10%-SIR 90% portfolios. The long/short portfolio alphas account for the overperformance or underperformance of long/short portfolios with respect to the Fama and French Five Factor Model benchmark.

We see the alphas go to zero during March 2006 and March 2009 accounting for the lack of performance in the long/short portfolios during these time periods. Generally, the SIR 5%-SIR 95% has the consistently higher alpha than the SIR 10%-SIR 90% portfolio across my study time period. The total range of the alphas for the long/short portfolios is between -0.5 and 3, this is a much smaller range than the heavily or lightly shorted portfolios shown in Figure 12. There is a good deal of correlation between the two portfolios, which is to be expected, since the SIR 5%-SIR 95% portfolio stocks will also be present in the SIR 10%-SIR 90% portfolio.

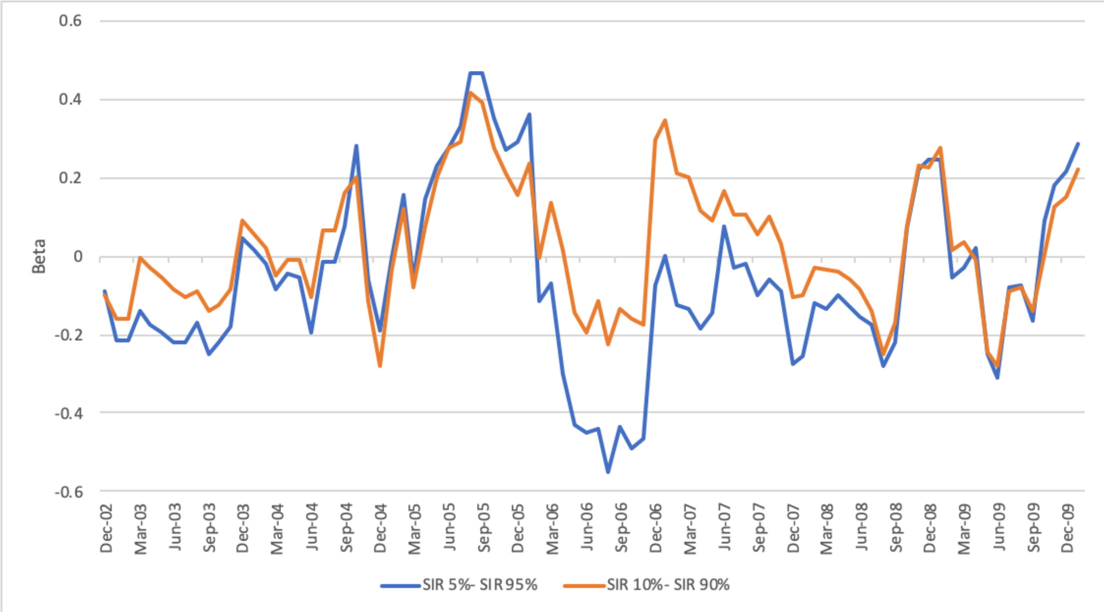
Figure 14: SIR Portfolio Rolling Regression Betas for the United States



Notes: Figure 14 shows the SIR Portfolio Rolling Regression Betas for the United States. The rolling regression is conducted across a rolling 24-month period from January 2001 to January 2010 using the Fama and French Five Factor Model. The dates shown in Figure 14 account for the end period of that particular rolling regression, which is conducted on a monthly basis. I report for Betas, which show the volatility of the portfolios with respect the market.

Figure 14 shows the rolling regression betas for the United States using the SIR 95%, SIR 90%, SIR 10% and SIR 5% portfolios. We see a good level of correlation with all 4 portfolio betas. We see a decrease in beta between December 2005 and December 2006 for all 4 portfolios, this decline could be due to other factors such as SMB, HML, CMA or RMW driving the returns process. The betas remain consistently between 0.9 and -1 for all portfolios, indicating no extreme fluctuations in portfolios with respect to the market. Betas higher than 1 or lower than -1 can be indicative of vast fluctuations with the market, often beneficial for speculators over investors. It is natural for the beta of these portfolios to vary, since the other factors of the Fama and French Five Factor Model may be contributing more the performance of portfolios over the beta factor.

Figure 15: Long/Short SIR Portfolio Rolling Regression Betas for the United States



Notes: Figure 15 shows the Long/Short SIR Portfolio Rolling Regression Betas for the United States. The rolling regression is conducted across a rolling 24-month period from January 2001 to January 2010 using the Fama and French Five Factor Model. The dates shown in Figure 15 account for the end period of that particular rolling regression, which is conducted on a monthly basis. I report for Betas, which show the volatility of the portfolios with respect the market.

Figure 15 shows the rolling regression betas for the United States using the SIR 5%-SIR 95% and SIR 10%-SIR 90% portfolios. The long/short portfolio betas account for the volatility of long/short portfolios with respect to the market. We see a good level of correlation with all 2 portfolio betas. The range of the betas vary between 0.5 and -0.6, indicating that these portfolios are much less volatile than the market portfolio and my standard high and low short interest portfolios. These portfolios can be held if market volatility risk is a concern for an investor. It is natural for the beta of these portfolios to vary, since the other factors of the Fama and French Five Factor Model may be contributing more the performance of portfolios over the beta factor.

We started off this research question with the following: "In adjusting for risk, Boehmer et al. (2010) uses the Fama and French (1993) three-factor model augmented by the momentum factor. However, is this long/short strategy still valid if a different and more recent model, such as Fama and French (2015) five-factor model, is used to adjust for risk premium?".

To show good news in short interest we would need a higher positive raw and excess return on lightly shorted portfolios compared to heavily shorted portfolios. We would also need a significantly higher alpha on my asset pricing model (in my case the Fama and French (2015) Five Factor Model) for my lightly shorted portfolios compared to heavily shorted portfolios.

Based on these two criteria, I can say that there is indeed good news to short interest when the Fama and French (2015) Five Factor Model is used to adjust for risk premium. Both SIR 5% and SIR 10% stock portfolios have a higher raw and excess return than the SIR 95% and SIR 90% stock portfolios. Both SIR

5% and SIR 10% portfolios have a significantly higher alpha compared to the SIR 95% and SIR 90% portfolios indicating you are achieving a much higher return for the level of undiversifiable risk you are taking. It must be noted that although Boehmer et al. (2010) employed a long/short strategy, I only advise a long SIR 5% portfolio strategy, due to the positive raw and excess returns on the SIR 95% and SIR 90% portfolios. Boehmer et al. (2010) found a slight negative raw return in their most shorted portfolio, however the return was so close to zero a shorting strategy would have to be revaluated in their case as well. Generally long-term short positions in the market need to be taken with a great degree of caution, as there are more average performing companies than there are poor performing companies.

6.2 Is There Arbitrage? (Research Question 2)

This research question aimed to answer the following question: After the publication of Boehmer et al. (2010), does the opportunity for excess returns remain, or have investors adopted this strategy and arbitraged away the excess returns? I first report the descriptive statistics of the individual portfolios formed before the regression for these portfolios is run in **Table 15**, the main results for the regression are reported in **Table 16**.

Table 15: Descriptive Statistics for Portfolios

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Portfolios	Mean	Standard Error	Median	Standard Deviation	Kurtosis	Skewness	Range	Maximum	Minimum
SIR 95%	0.017	0.007	0.021	0.064	0.083	-0.380	0.307	0.162	-0.145
SIR 90%	0.016	0.006	0.020	0.061	0.257	-0.263	0.311	0.175	-0.136
SIR 5%	0.026	0.004	0.028	0.043	-0.097	-0.370	0.199	0.107	-0.092
SIR 10%	0.021	0.004	0.023	0.040	0.235	-0.368	0.207	0.116	-0.091
SIR 5%-SIR95%	0.009	0.004	0.010	0.036	-0.400	0.234	0.159	0.097	-0.061
SIR 10%-SIR90%	0.005	0.003	0.003	0.030	-0.211	0.314	0.142	0.082	-0.060

Notes: Data source using Microsoft Excel from Thomson Reuters US Total Return Index DataStream Data. Table 15 shows the descriptive statistics for the individual portfolios formed before regression including the standard and long/short portfolios. I report the mean, standard error, median, standard deviation, sample variance, kurtosis, skewness, range, minimum and maximum values of the portfolios. The descriptive statistics above give the distribution of the return of the portfolios on a total return basis. Data author's own calculation.

Table 15 shows the descriptive statistics for portfolios run in the regression. Looking at Table 15 in Column (1) we notice the mean is higher in the lightly shorted portfolios over the heavily shorted portfolios and this indicates the better performance of the lightly shorted stocks over the heavily shorted stocks. Both heavily and lightly shorted portfolios show an acceptable level of standard error shown in Column (2) and thus we can be confident in my forecasting abilities using these portfolios. The median shown in Column (3) is higher in both highly and lightly shorted portfolios and indicates negative skew in the data, which can also be seen from the skewness figures.

Regarding standard deviation shown in Column (4), we see a much higher standard deviation in heavily shorted portfolios as opposed to lightly shorted portfolios. This indicates a much higher volatility in heavily shorted portfolios, which was very evident when constructing the dataset. This is very much similar to my results in research question 1 and emphasises that heavily shorted portfolios are much more volatile than their lightly shorted counterparts. This could be a result of short squeezes amplifying the upside and also speculators shorting on the downside. The stocks with high short interest are more likely to interest speculators over investors, since volatility is beneficial to the speculator as it creates low prices to buy and high prices to sell. However, volatility can be a disadvantage to the investor if he or she is forced out of his or her position because of fear of irreversible losses in the market.

The results for the Kurtosis shown in Column (5) are much more interesting since there is no defined trend in whether heavily or lightly shorted portfolios have fat tails in this case. The skewness in Column (6) indicates a negative skew in all portfolios favouring more months of portfolios holding higher returns than lower returns, which would make sense since the purpose for holding stocks is to gain a return rather than make a loss.

The range in Column (7) indicates a greater variation in returns for the heavily shorted portfolios over the lightly shorted portfolios, this matches up with the higher standard deviation the heavily shorted portfolios exhibit. This effect can also be seen in Columns (8) and (9) for the maximum and minimum returns for a particular month for each portfolio.

Table 16: Regression Analysis Results of Monthly Equal Weighted Returns on Highly and Lightly Shorted Stock Portfolios for the United States

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Portfolios/#Stocks	Raw Return	Excess Return	Intercept	EM-RF	SMB	HML	MOM
SIR 95%	0.017	0.017	0.003	1.203***	1.105***	0.104	-0.280***
#114 stocks			(0.219)	(0.000)	(0.000)	(0.316)	(0.000)
SIR 90%	0.016	0.016	0.002	1.179***	1.083***	0.069	-0.242***
#227 stocks			(0.249)	(0.000)	(0.000)	(0.355)	(0.000)
SIR 5%	0.026	0.026	0.015***	0.928***	0.281***	0.257***	-0.027
#114 stocks			(0.000)	(0.000)	(0.005)	(0.008)	(0.708)
SIR 10%	0.021	0.021	0.010***	0.949***	0.238***	0.204***	-0.006
#227 stocks			(0.000)	(0.000)	(0.000)	(0.000)	(0.896)
SIR 5%-SIR 95%	0.009	0.009	0.012***	-0.275***	-0.824***	0.153	0.254**
			(0.000)	(0.002)	(0.000)	(0.259)	(0.014)
SIR 10%-SIR 90%	0.005	0.005	0.008***	-0.230***	-0.846***	0.136	0.237***
			(0.000)	(0.000)	(0.000)	(0.138)	(0.000)

Notes: Table 16 shows the regression analysis of monthly equal weighted highly and lightly shorted stock portfolios. Data source from Microsoft Excel from Thomson Reuters US Total Return Index DataStream Data. The regression equation[68] run to obtain each portfolio in the table is as follows:

$$R_{port\,t} - R_{ft} = \alpha_{port} + \beta_{port}\left(R_m - R_f\right)_t + \gamma_{port}SMB_t + \delta_{port}HML_t + \eta_{port}MOM_t + \varepsilon_t \quad (49)$$

The dependent variable is the equal weighted monthly portfolio excess return $R_{port\,t} - R_{ft}$ on highly or lightly shorted stock portfolios for the month subsequent to portfolio creation on stocks from the Thomson Reuters US Total Return Index. $\left(R_m - R_f\right)_t$ is the market risk premium, SMB_t is the excess return of a portfolio of small cap stocks over a portfolio of large cap stocks, HML_t is the excess return of a portfolio of high book to market stocks over low book to market stocks and MOM_t[69] is the excess return of period t-1 winners portfolio over t-1 losers portfolio. The monthly data for the three Fama French factors and the one Momentum factor is from Kenneth French's Data Library. The portfolios SIR 95% and SIR 90% include stocks with short interest ratios from the 95th and 90th percentile respectively in month t-1. The portfolios SIR 5% and SIR 10% include stocks with short interest ratios from the 5th and 10th percentile respectively in month t-1. The portfolios are rebalanced each month based on new short interest data. Portfolio raw returns and excess returns over the risk-free rate are shown under the headings of Raw Return and Excess Return. The number of stocks in each portfolio can change each month and thus the average number of stocks in each portfolio is shown under the heading # stocks. I report p values in brackets under respective regression coefficients. Stars are used to indicate significant figures. * is significance at the 10% level, ** is significance at the 5% level and *** is significance at the 1% level. My sample period is based on monthly short interest data from February 2010 to July 2017. Data author's own calculation.

I next report the regression analysis results in **Table 16**. If we look at Column (1) and (2) of Table 16, we can see that the SIR 90% portfolio has the lowest raw return of 1.6% a month and the SIR 5% portfolio has the highest raw return of 2.6% a month, this is a significant difference in returns and shows that on average lightly shorted stocks do indeed outperform heavily shorted stocks. The outperformance of lightly shorted stocks over heavily shorted stocks is consistent with Boehmer et al. (2010), Desai et al. (2002) and Asquith

68 It has been advised that running a regression with the CME factor on top is of interest as well, i.e. regressing with the Dreschler and Dreschler (2015) five factor model. The issue arises with obtaining the factor loadings of CME for my datasets, though in further study this model could be implemented to see the effect of the CME factor.

69 I also regress equation (49) without MOM to facilitate a comparison between the two periods of 2001-2010 and 2010-2017. I find very little difference in R Square and good confidence for EF-RM, HML and SMB factors. I believe that the contribution of MOM is minimal to the 2010-2017 period in driving returns.

et al. (2005).

This performance of lightly shorted stocks over heavily shorted stocks is relatively consistent across the dataset, we can see in Column (1) that the SIR 10% portfolio has a raw return of 2.1% a month and the SIR 95% portfolio has a raw return of 1.7% a month. However, the SIR 95% portfolio does outperform the SIR 90% portfolio, so it is not a completely even distribution in return. What we can see is however is that all the lightly shorted stock portfolios outperform the heavily shorted stock portfolios.

The excess returns are very much similar to the raw returns for standard portfolios due to our very low interest rates imposed by central banks. After the 2007-2009 financial crisis, interest rates were cut to all-time lows [70] to stimulate the economy and this has meant that the risk-free rate has been marginally above zero. Much more of a difference in excess returns was noted by Boehmer et al. (2010) since central banks provided more normalised interest rates.

Next looking at the results of the regression, we see that in Column (3) the intercepts range from 0% for the SIR 95% portfolio to 0.2% for the SIR 90% portfolio. The intercepts indicate the return that the model has failed to specify, i.e. the alpha of the model. Again, positive intercept coefficients indicate that the return data has performed better than the model expected, negative intercept coefficients indicate that the return data has performed worse than the model expected. Thus, highly shorted portfolios for February 2010 to July 2017 seem to perform in line with the benchmark, which is in slight contradiction to the findings of Boehmer et al. (2010). This over performance may very well be due to the fact we have been in a bull market for this time period and stocks in general have gone up more than they have gone down. It is worth noting that the highly shorted portfolios perform worse on an alpha basis than the entire dataset and in particular the lightly shorted portfolios. The SIR 5% portfolio has a particularly strong alpha of 1.5% a month and this remains the best portfolio to hold on a raw return and alpha basis.

Looking further into the results at Column (4) we can see that the heavily shorted portfolios have larger market betas than their lightly shorted counterparts, we also see in Column (5) that the heavily shorted portfolios have a positive coefficient on SMB and in Column (6) a zero coefficient on HML. Thus, I can say that these stocks resemble small cap growth stocks judging by my regression results. This was also noted by Boehmer et al. (2010) and Desai et al. (2002). The heavily shorted portfolios also have a significant negative momentum loading compared to lightly shorted portfolios, which can be seen in Column (7).

The larger market betas indicate that highly shorted portfolios have a greater degree of market risk and thus in the past have been more volatile than the market as a whole. This was evident when forming the highly shorted portfolios, these portfolios were much more susceptible to large decreases and subsequently large increases. The stocks in these portfolios would hugely favour day traders as increased volatility means a greater chance of turning a profit in a small period of time. The volatility may be a case of a lot of traders

---

70 The Federal Reserve Discount Rate was cut to 0.25%, the Bank of England Base Rate was cut to 0.5% and the Bank of Japan Base Rate was cut to 0.3%. Interest rates have struggled to move since then and now in September 2019 the Bank of England Base Rate is at 0.75%, the Bank of Japan Base Rate is at -0.1%, while the Federal Reserve Discount Rate is at 2.25%. Central Banks have stated that weak wage growth and low inflation have hindered the rise of interest rates to more normal levels.

suddenly moving in and out of these stocks due to short squeezes and in particular thin order books[71] as seen in Ahn et al. (2001).

Moving onto the long/short portfolios we see that in Column (3) the monthly abnormal return of the SIR 5%-SIR 95% is at 1.2% per month. The portfolio also holds a negative market beta of -0.275 of significance, as shown in Column (4). This means that if this portfolio was to be added to an existing holding, returns would be boosted while at the same time market risk being reduced. Boehmer et al. (2010) proposed this strategy of going long the least shorted stock portfolio and short the most shorted stock portfolio, this can be modelled with the SIR 5%-SIR 95% portfolio. This was proposed because of the negative beta and the large abnormal return, I however do not propose this. It is again worth noting that the long/short portfolios and long only portfolio cannot be compared directly due to the cost of shorting associated with borrowed stocks. However even if we consider the cost of borrowing as seen in Lu et al. (2018), the long/short portfolios greatly underperform the long only portfolios. This is due to the high cost of borrowing associated with some highly shorted stocks.

In my findings, the best performing portfolio by far has been the SIR 5% portfolio and it seems most viable to hold the SIR 5% portfolio and not apply a short strategy on the other end.

Table 17: Regression Statistics for Portfolios

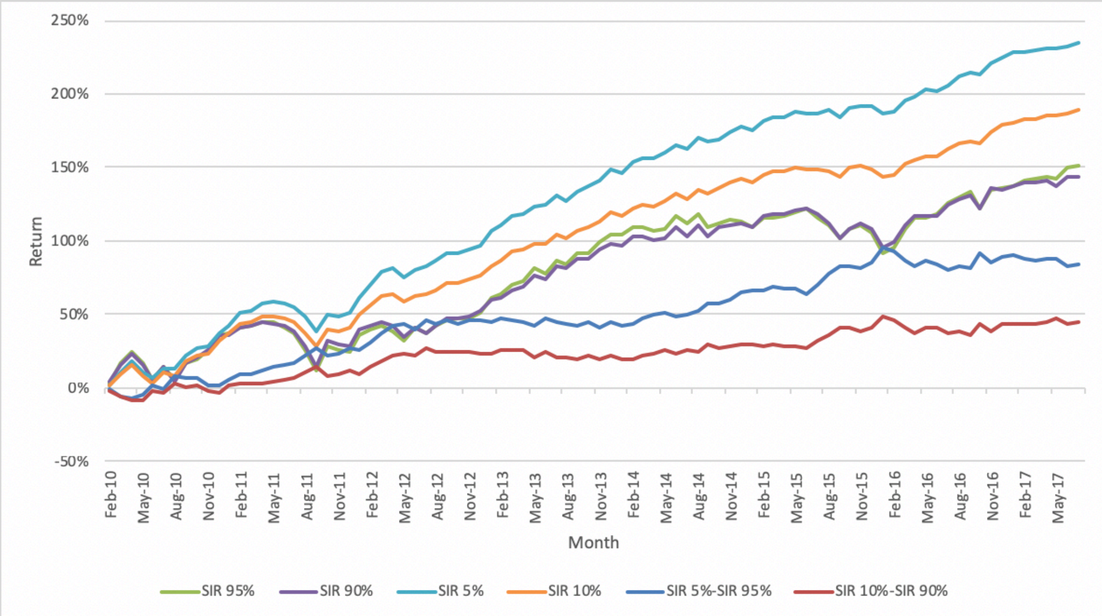
	(1)	(2)	(3)	(4)
Portfolios	Multiple R	R Square	Adjusted R Square	Observations
SIR 95%	0.950	0.902	0.898	90
SIR 90%	0.972	0.945	0.943	90
SIR 5%	0.904	0.817	0.808	90
SIR 10%	0.959	0.919	0.915	90
SIR 5% - SIR 95%	0.697	0.486	0.461	90
SIR 10% - SIR 90%	0.813	0.660	0.644	90

Notes: Table 17 gives the regression statistics for each individual portfolio. Again, we list statistics for Multiple R, R Square, Adjusted R Square and Observations. Data author's own calculation.

Table 17 shows the regression statistics for portfolios. Looking at Table 17 in Column (4) we see that we have 90 observations for the months of February 2010 to July 2017 inclusive. In Column (2) we see a stronger R Square in my larger portfolios of SIR 90% and SIR 10%. Overall, we see a strong R Square across all standard portfolios showing a good link of this regression model with the data. In regards to comparison of fit with research question 1, we see a much stronger fit with the Carhart Four Factor Model over the Fama and French Five Factor Model.

[71] The order book shows the liquidity in a stock, when an order book is thin it lacks the limit orders (market depth) needed to maintain relatively stable prices, this can lead to large jumps in prices, i.e. the stock gapping. Market makers are often employed to increase liquidity in illiquid stocks, their job is to make sure there is a buyer available for every seller and a seller available for every buyer. The NYSE in particular has maintained designated market makers to keep liquidity up, spreads smaller and prices more stable and efficient.

Figure 16: Cumulative SIR Portfolio Returns for the United States



Notes: Data Source using Microsoft Excel from Thomson Reuters US Total Return Index DataStream Data. Figure 16 shows the performance of different SIR portfolios over my study period.

Figure 16 shows the cumulative returns [72] of the different short interest ratio (SIR) portfolios over the period of the study. Looking at Figure 16 we can see that there is a good deal of positive correlation between the standard portfolios, in particular two dips in August 2011 and December 2015 stand out. We also see that the long/short portfolios show good positive correlation with each other. We can easily see that overtime the outperformance of the SIR 5% portfolio is evident and we can see that the SIR 5% portfolio outperforms the SIR 95% portfolio by a margin over 90%. The SIR 5% performance drastically outperforms the others, including the SIR 10% portfolio which has stocks that are included in the SIR 5% portfolio. The worst performing portfolio is the SIR10%-SIR90% portfolio which for a period of two years underperformed the market as a whole. This again highlights the negative of employing a shorting strategy based on short interest levels, it is evident that the best performance comes from lightly shorted stocks, but the performance of heavily shorted stocks is not bad enough to warrant a shorting strategy.

[72] Cumulative being the return that would be achieved if the portfolio were bought at the start of February 2010 and held to a particular date in the future. This allows us to see how the different portfolios developed over time and whether they have any correlation between each other. The portfolios are of course on a total return basis and include dividends reinvested at the first available opportunity of market open.

Table 18: Correlation Matrix for SIR Portfolio Returns for the United States

Portfolios	SIR 95%	SIR 90%	SIR 5%	SIR 10%	SIR 5%-SIR 95%	SIR 10%-SIR 90%
SIR 95%	1	0.986	0.843	0.883	-0.772	-0.815
SIR 90%	0.986	1	0.857	0.901	-0.731	-0.819
SIR 5%	0.843	0.857	1	0.977	-0.309	-0.430
SIR 10%	0.883	0.901	0.977	1	-0.406	-0.488
SIR 5%-SIR 95%	-0.772	-0.731	-0.309	-0.406	1	0.931
SIR 10%-SIR 90%	-0.815	-0.819	-0.430	-0.488	0.931	1

Notes: Table 18 shows the Correlation Matrix for SIR Portfolio Returns for the United States. This correlation matrix matches the return correlations between high, low and long/short SIR portfolios. A correlation of 1 indicates perfect positive correlation, a correlation of -1 indicates perfect negative correlation and a correlation of 0 indicates little to no correlation.

Table 18 shows the correlation matrix for SIR portfolio returns for the United States. The correlation matrix allows us to see how correlated returns are between different portfolios. We see the strongest positive correlation between the SIR 90% and SIR 95% portfolios of 0.986; the SIR 5% and SIR 10% portfolios also hold a strong positive correlation of 0.977. This is to be expected since the SIR 10% portfolio includes stocks from the SIR 5% portfolio and the SIR 90% portfolio includes stocks from the SIR 95% portfolio. Even with this knowledge, the correlations remain positively strong and again show returns may very well be influenced from stocks in the extreme percentiles.

The strongest negative correlation of -0.819 is from the SIR 90% portfolio and SIR 10%-SIR 90% portfolio. This is again due to the short component of the later portfolio implementing a perfect negative correlation. The SIR 10% long component of the long/short portfolio means we do not have a perfect negative correlation, but a strong negative correlation.

Regarding arbitrage, the Chow Breakpoint Test can be used to check whether there is any statistically significant structural break (known or unknown) exists in the data. This is only however applicable if the models used in both periods (2001 to 2010) and (2010 to 2017) are identical. However, this test is not completely necessary since my findings for research question 2 show that arbitrage has not taken place and thus a structural break (of significance) does not exist in the data. A rejection of my null hypothesis of a structural break would need to be tested using the Chow Breakpoint Test to see the significance of the structural break. However, we will run the Chow Breakpoint Test to illustrate this point using the Fama and French Three Factor Model with Momentum in both periods of 2001 to 2010 and 2010 to 2017 for high and low short interest portfolios.

Suppose we were to model portfolios in my entire time period dataset from 2001 to 2017 using the Fama and French Three Factor Model with Momentum.

$$R_{port\,t} - R_{ft} = \alpha_{port} + \beta_{port}\left(R_m - R_f\right)_t + \gamma_{port}SMB_t + \delta_{port}HML_t + \eta_{port}MOM_t + \varepsilon_t \quad (50)$$

If we were then to split my portfolio data into two groups, one for 2001 to 2010 and another one for 2010 to 2017, we would have:

$$R_{port\,t} - R_{ft} = \alpha_1 + \beta_1 \left(R_m - R_f \right)_t + \gamma_1 SMB_t + \delta_1 HML_t + \eta_1 MOM_t + \varepsilon_t \tag{51}$$

And

$$R_{port\,t} - R_{ft} = \alpha_2 + \beta_2 \left(R_m - R_f \right)_t + \gamma_2 SMB_t + \delta_2 HML_t + \eta_2 MOM_t + \varepsilon_t \tag{52}$$

The null hypothesis of the Chow Breakpoint Test will say that $\alpha_1 = \alpha_2$, $\beta_1 = \beta_2, \gamma_1 = \gamma_2$, $\delta_1 = \delta_2$ and $\eta_1 = \eta_2$. We also hold the assumption the model errors are iid (independent and identically distributed) with a variance that is unknown. Using this information, the Chow Breakpoint Test Statistic is defined as follows:

$$Chow = \frac{(RSS - RSS_1 - RSS_2)/k}{(RSS_1 + RSS_2)/(n_1 + n_2 - 2k)} \sim F_{k, n_1 + n_2 - 2k} \tag{53}$$

Where RSS is the sum of squared residuals of the combined time period of 2001 to 2017 for a portfolio, RSS_1 is the sum of squared residuals of the time period 2001 to 2010 for the same portfolio and RSS_2 is the sum of squared residuals of the time period 2010 to 2017 for the same portfolio. n_1 and n_2 are the number of observations in each group and k is the total number of parameters, in my case 5. The test follows the F distribution with k and $n_1 + n_2 - 2k$ degrees of freedom.

Table 19: Chow Breakpoint Test for Portfolios using Fama and French Three Factor Model with Momentum over Periods 2001-2010 and 2010-2017

	(1)	(2)
Portfolios	Chow Statistic	P Value
SIR 95%	8.275***	(0.000)
SIR 90%	8.410***	(0.000)
SIR 5%	4.830***	(0.000)
SIR 10%	7.134***	(0.000)
SIR 5% - SIR 95%	4.546***	(0.001)
SIR 10% - SIR 90%	5.975***	(0.000)

Notes: Table 19 shows the Chow Breakpoint Test for Portfolios using the Fama and French Three Factor Model with Momentum. The test is conducted over periods 2001-2010 and 2010-2017 for SIR 95%, SIR 90%, SIR 5%, SIR 10%, SIR 5%-SIR95% and SIR 10%-SIR90% portfolios using monthly returns data. The Chow Statistic is calculated using the following formula for each portfolio:

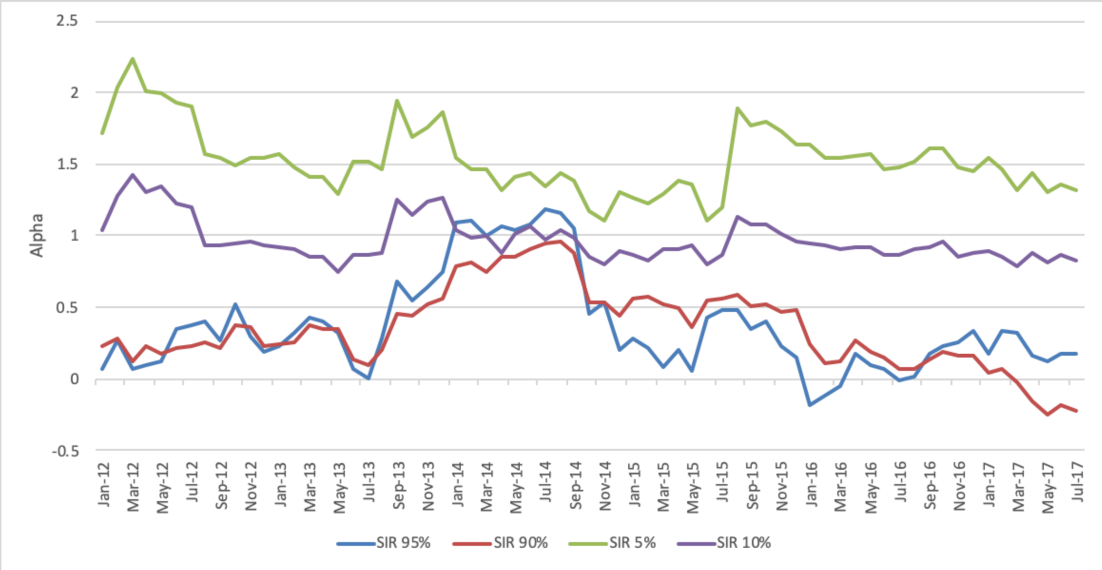
$$Chow = \frac{(RSS - RSS_1 - RSS_2)/k}{(RSS_1 + RSS_2)/(n_1 + n_2 - 2k)} \sim F_{k, n_1 + n_2 - 2k} \tag{54}$$

Chow Statistic is tested using the F Test with 5 degrees of freedom in the numerator and 189 degrees of freedom in the denominator. This is obtained using 5 parameters for k, 90 observations for n_1 and 109 observations for n_2. This constitutes a total of 199 months for each portfolio. P values are denoted in brackets. I use a star system to denote significance, *** is significance at the 1% level, ** is significance at the 5% level and * is significance at the 10% level.

The Chow Breakpoint Test would be conducted for all portfolios and the results can be seen in **Table 19**. From the P-Values in the portfolios in the Chow Breakpoint Test we can see there is no statistically significant break between the two sets of data, further reinforcing my findings. In particular every portfolio has a significant Chow Statistic at the 1% confidence level denoted with ***.

Over my sample period, equity markets were not stable and experienced a variety of changes. Thus, in extension of my main results, I attempt to see if my findings are period specific or whether there are significant changes over time. In order to achieve this, I conduct a 24-month rolling regression using the Fama and French Three Factor Model with Momentum shown previously in Equation (49) to report for both alpha coefficient (abnormal returns) and beta coefficient (volatility with respect to market) across my sample period. I conduct the rolling regression for SIR 95%, SIR 90%, SIR 5%, SIR 10%, SIR 5%-SIR 95% and SIR 10%-SIR 90% portfolios, this means I account for heavily shorted portfolios, lightly shorted portfolios and long/short portfolios.

Figure 17: SIR Portfolio Rolling Regression Alphas for the United States



Notes: Figure 17 shows the SIR Portfolio Rolling Regression Alphas for the United States. The rolling regression is conducted across a rolling 24-month period from February 2010 to July 2017 using the Fama and French Three Factor Model with Momentum. The dates shown in Figure 17 account for the end period of that particular rolling regression, which is conducted on a monthly basis. I report for Alphas, which show the outperformance or underperformance of the portfolios with respect to the Fama and French Three Factor Model with Momentum benchmark.

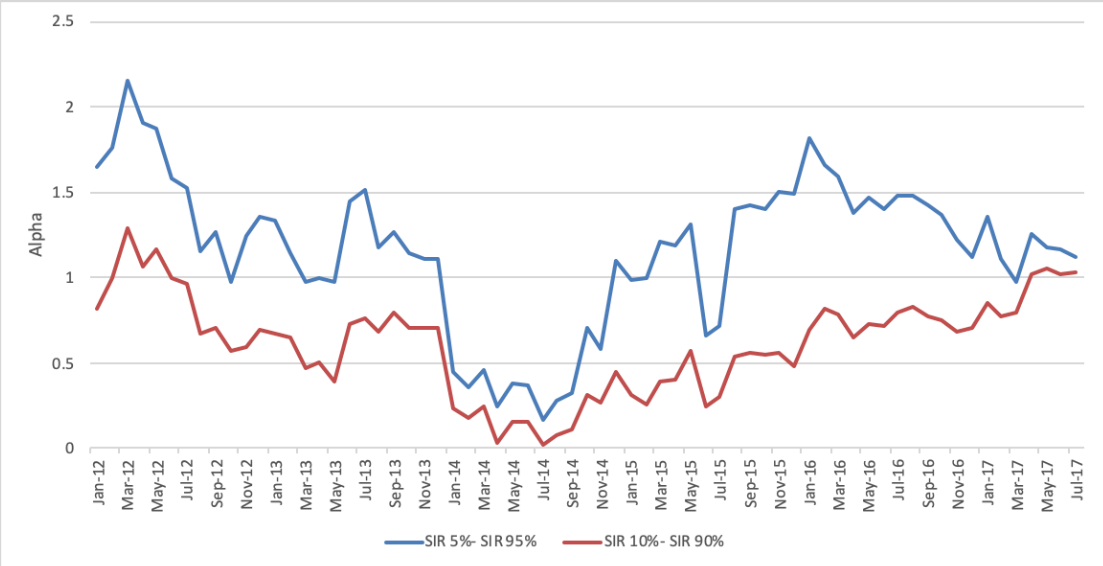
Figure 17 shows the rolling regression alphas for the United States using the SIR 95%, SIR 90%, SIR 10% and SIR 5% portfolios. We see that over this study period, lightly shorted portfolio alphas outperform heavily shorted portfolio alphas over the entire period in general. There is one exception where in July 2014, the convergence of alphas between the heavily and lightly shorted portfolios occurs. The reasoning behind this could be because of an added level of optimism in the market, leading to investors and speculators holding more long positions that usual.

Investors often become risk seeking when they should be risk adverse and risk adverse when they should be risk seeking in terms of asset allocation as noted in Lippi et al. (2018). Investors also tend to look at past performance in terms of asset allocation rather than being forward looking which they should be, often stocks can only increase or decrease so much before valuation becomes an issue. There is often a lack of liquidity to sustain already overvalued stock prices. A good example is the tech bubble of 1998-2000 where speculators and hedge funds kept pushing prices higher and higher from looking at performances of previous years as seen in Griffin et al. (2011). This could have led to a good performance in heavily shorted stocks as these stocks received less short interest than usual. These are common signs of an overheated market towards the end of 2014, where valuation fails to reflect future earnings growth potential. As stock prices rise, future returns diminish if earnings do not grow as fast as the market expects. Anything which affects earnings per shares in stocks (such as a recession) has a negative effect on stock prices, as stock prices reflect future earnings growth.

More recently towards July 2017, we can see that the alphas of the heavily shorted portfolios are turning negative, again highlighting their vast underperformance to lightly shorted portfolios. Also, we can see that more short positions are being formed in anticipation of an overdue bear market, which has not occurred since 2009.

The total range of alphas for the heavily shorted portfolios is between -0.3 and 1.2, the total range of alphas for the lightly shorted portfolios is between 0.8 and 2.2. Again, the range is similar between portfolios, but that lightly shorted portfolios hold the higher mean and median rolling regression alpha.

Figure 18: Long/Short SIR Portfolio Rolling Regression Alphas for the United States

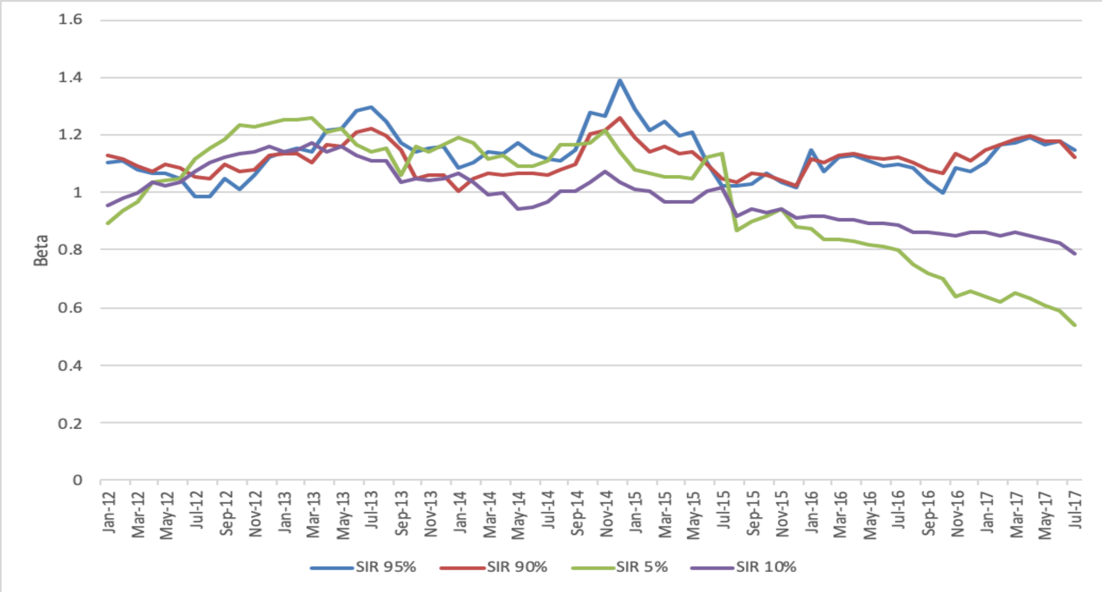


Notes: Figure 18 shows the Long/Short SIR Portfolio Rolling Regression Alphas for the United States. The rolling regression is conducted across a rolling 24-month period from February 2010 to July 2017 using the Fama and French Three Factor Model with Momentum. The dates shown in Figure 18 account for the end period of that particular rolling regression which is conducted on a monthly basis. I report for Alphas, which show the outperformance or underperformance of the portfolios with respect to the Fama and French Three Factor Model with Momentum benchmark.

Figure 18 shows the rolling regression alphas for the United States using the SIR 5%-SIR 95% and SIR 10%-SIR 90% portfolios. The long/short portfolio alphas account for the overperformance or underperformance of long/short portfolios with respect to the Fama and French Three Factor Model with Momentum benchmark.

We see that the SIR 5%-SIR 95% portfolio outperforms the SIR 10%-SIR 90% portfolio in terms of alpha over the study period. In July 2014, the outperformance is minimal and so this is the case in July 2017. Overall there is no point where the SIR 5%-SIR 95% portfolio has a lower alpha than the SIR 10%-SIR 90% portfolio. The total range of alphas for the long/short portfolios is between 0 and 2.1. There is a good deal of correlation between the portfolios, which is natural since the stocks composing the SIR 5%-SIR 95% portfolio will also be in the SIR 10%- SIR 90% portfolios.

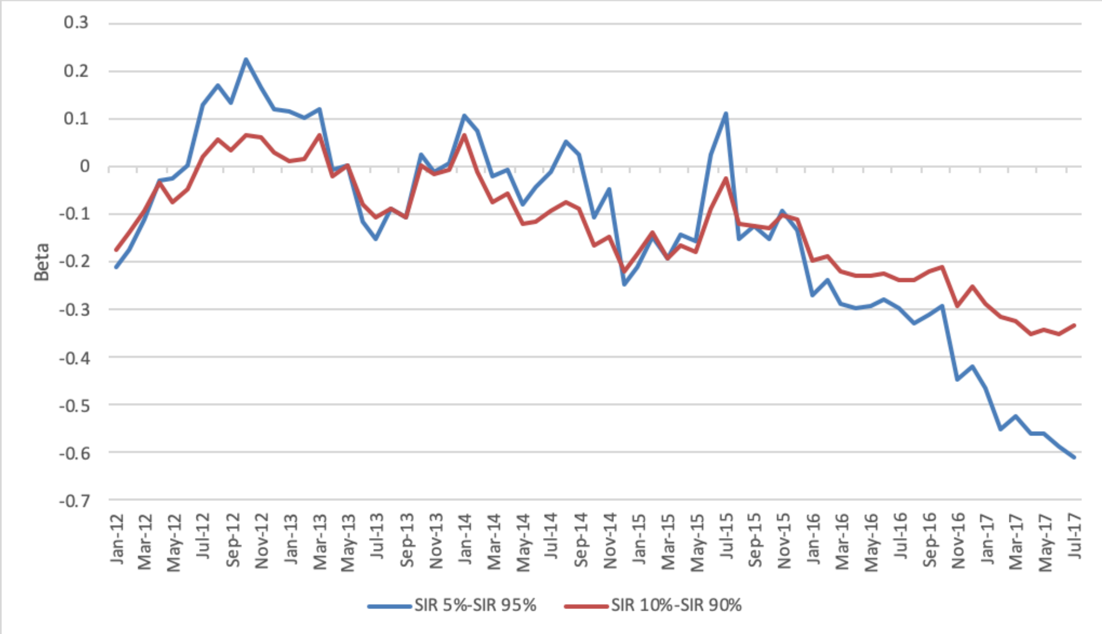
Figure 19: SIR Portfolio Rolling Regression Betas for the United States



Notes: Figure 19 shows the SIR Portfolio Rolling Regression Betas for the United States. The rolling regression is conducted across a rolling 24-month period from February 2010 to July 2017 using the Fama and French Three Factor Model with Momentum. The dates shown in Figure 19 account for the end period of that particular rolling regression, which is conducted on a monthly basis. I report for Betas, which show the volatility of the portfolios with respect the market.

Figure 19 shows the rolling regression betas for the United States using the SIR 95%, SIR 90%, SIR 10% and SIR 5% portfolios. We see a good level of correlation with all 4 portfolio betas, until about May 2015, where there is a clear divergence in the betas of the lightly shorted portfolios over the betas of the heavily shorted portfolios. This divergence indicates that lightly shorted portfolios are becoming less volatile than the market as a whole compared to heavily shorted portfolios. It is natural for the beta of these portfolios to vary, since the other factors of the Fama and French Three Factor Model with Momentum could be contributing more the performance of portfolios over the beta factor.

Figure 20: Long/Short SIR Portfolio Rolling Regression Betas for the United States



Notes: Figure 20 shows the Long/Short SIR Portfolio Rolling Regression Betas for the United States. The rolling regression is conducted across a rolling 24-month period from February 2010 to July 2017 using the Fama and French Three Factor Model with Momentum. The dates shown in Figure 20 account for the end period of that particular rolling regression which is conducted on a monthly basis. I report for Betas, which show the volatility of the portfolios with respect the market.

Figure 20 shows the rolling regression betas for the United States using the SIR 5%-SIR 95% and SIR 10%-SIR 90% portfolios. The long/short portfolio betas account for the volatility of long/short portfolios with respect to the market. We see a good deal of correlation between the 2 portfolios, however after May 2015 we see a noticeable drop in the betas of both portfolios. This is again linked to the fact that the betas of the lightly shorted portfolios have been dropping compared to the betas of the heavily shorted portfolios. It is natural for the beta of these portfolios to vary, since the other factors of the Fama and French Three Factor Model with Momentum may be contributing more the performance of portfolios over the beta factor.

Therefore concluding, my initial problem was the following: After the publication of Boehmer et al. (2010), does the opportunity for excess returns remain, or have investors adopted this strategy and arbitraged away the excess returns? My results show that that low short interest stocks still outperform high short interest stocks, however the strategy proposed by Boehmer et al. (2010) is not valid due to the positive performance of the SIR 95% portfolio. If the SIR 95% portfolio lead to a negative return [73] then shorting it would be a viable strategy, however this is not the case.

The best strategy to employ would be to go long the SIR 5% portfolio and to not hold a short portfolio. By doing this we would have achieved a raw return of 2.6% a month and a risk adjusted alpha of 1.5% a month. So, my findings show that the opportunity for excess returns does remain, but the nature of the bull market

73 Boehmer et al. (2010) found that the SIR 99% portfolio with equal weighting had an average monthly return of -0.1%, which would indeed have been a negative return and a viable shorting strategy.

has meant that following the strategy of Boehmer et al. (2010) has not been viable for this time period.

My results show that arbitrage has not taken place, as the performance of the SIR 95% portfolio and the SIR 5% portfolio would have to be similar, the SIR 5% portfolio clearly outperforms the SIR 95% portfolio on both a raw return basis and a risk adjusted alpha basis. For this strategy to be arbitraged I believe more market participants would be required outside the hedge funds, the difficult nature of this strategy implementation may mean the large volumes needed to show arbitrage do not exist. Given that Boehmer et al. (2010) showed this strategy in 2010 in an academic journal may also affect the audience. It is worth noting why hedge funds have not implemented this strategy and this remains one of the most interesting findings of this thesis. It is well known that hedge funds will try strategies to generate money that ordinary investors would not consider, especially strategies that involve algorithmic trading (trading with computer algorithms) and high frequency trading (trading in very short periods of time). One possible explanation for this could be the generation of excess returns may not be great enough compared to other strategies (like holding good companies based on fundamental analysis) or that the borrowing costs[74] on the short end may be too much for some smaller hedge funds to rebalance with as pointed out in Lu et al. (2018). There is a cost with borrowing costs and hedge funds need to take this into account when evaluating asset return strategies.

Overall the implementation of a long SIR 5% portfolio strategy is still somewhat difficult since the composition of the SIR 5% portfolio will change every month. A portfolio of 104 stocks is enough to remove [75] most of the unsystematic risk in relation to a benchmark such as the S&P 500. However, it is very well possible that an investor would find themselves selling 104 stocks and buying 104 stocks each month, which makes the strategy somewhat tedious and expensive. An algorithmically driven hedge fund is much more suited to this strategy and I believe that is where the benefit lies. There is also the potential for investors to gain exposure to such strategy via an exchange traded fund (ETF) that focuses on a high/low short interest monthly rebalancing strategy. This concludes my findings for research question 2.

6.3 Is the Strategy Valid in another OECD country that of Canada ? (Research Question 3)

This research question aimed to answer the following question: Is the long/short strategy of Boehmer et al. (2010) still valid in an international OECD market that of Canada? It is worth noting that accessing returns data is not much of an issue across different markets, however short interest data is much more difficult to find in large quantities, especially across a time period stretching across several years. This can cause results of smaller datasets such as Canada to experience more volatility as a whole, due to less observations in each portfolio. However, I have tried my best to include as many stocks as possible for each country in order to

[74] For example the borrowing costs for stocks have been as high as 70% of share value for some stocks with the highest short interest. It is common for borrowing costs to be over 10% of the price of a stock in stocks with high levels of short interest (Bary, 2019).

[75] As the number of securities in a portfolio increases, the more and more the unsystematic risk of the portfolio is removed. To get the unsystematic risk completely removed an investor would have to hold the market portfolio, however just by holding 10-12 stocks of the S&P 500 an investor can remove approximately 95% of the unsystematic risk of the S&P 500. Any additional stocks will not make too much of a difference to the level of unsystematic risk and may impact the alpha of the portfolio.

increase the robustness of my results. From the start, I knew I would not be able to replicate the depth of the US market due to generally less listings and less information transparency. However, I believe that achieving a dataset with 280 firms per month on average is very much statistically significant and should provide a good insight into this strategy outlined by Boehmer et al. (2010) in the Canadian market. I first report the descriptive statistics of the individual portfolios formed before the regression for these portfolios is run in **Table 20**, the main results for the regression are reported in **Table 21**.

Table 20: Descriptive Statistics for Portfolios

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Portfolios	Mean	Standard Error	Median	Standard Deviation	Kurtosis	Skewness	Range	Maximum	Minimum
SIR 95%	0.010	0.005	0.011	0.051	0.455	0.326	0.271	0.161	-0.110
SIR 90%	0.009	0.005	0.008	0.047	0.010	0.265	0.232	0.139	-0.093
SIR 5%	0.025	0.005	0.019	0.048	7.982	1.678	0.361	0.243	-0.119
SIR 10%	0.021	0.004	0.020	0.035	1.927	0.383	0.220	0.138	-0.082
SIR 5%-SIR95%	0.015	0.006	0.013	0.054	3.594	0.935	0.349	0.230	-0.119
SIR 10%-SIR90%	0.012	0.004	0.005	0.039	0.044	0.472	0.185	0.120	-0.066

Notes: Data source using Microsoft Excel from Thomson Reuters Canada Total Return Index DataStream Data. Table 20 shows the descriptive statistics for the individual portfolios formed before regression including the standard and long/short portfolios. I report the mean, standard error, median, standard deviation, sample variance, kurtosis, skewness, range, minimum and maximum values of the portfolios. The descriptive statistics above gives the distribution of the return of the portfolios on a total return basis. Data author's own calculation.

Table 20 shows the descriptive statistics for portfolios run in the regression. Looking at Table 20 in Column (1) on a mean return basis, we see that the lightly shorted portfolios once again outperform the heavily shorted portfolios. The standard error shown in Column (2) for both heavily and lightly shorted portfolios is again at a reasonable level to aid forecasting. In most cases we see in Column (3) that the median is below the mean indicating skew in the data.

Regarding the standard deviation shown in Column (4), we see a slightly overall higher standard deviation in the heavily shorted portfolios, however as I have once said before, as my sample dataset becomes smaller and smaller we risk more firm specific risk into my dataset rather than short interest characteristics.

We do see an extreme level of Kurtosis in Column (5) in my SIR 5% portfolio indicating there may be an outlier or outliers in the portfolio that is performing much better than the other stocks in the portfolio. This is also confirmed with the skewness data, showing an extreme return in the SIR 5% portfolio. The range of all portfolios shown in Column (7) is relatively similar between 0.185 and 0.361 and so is the level of minimum and maximum levels shown in Columns (8) and (9) respectively.

Table 21: Regression Analysis Results of Monthly Equal Weighted Returns on Highly and Lightly Shorted Stock Portfolios for Canada

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Portfolios/ #Stocks	Raw Return	Excess Return	Intercept	EM-RF	SMB	HML	MOM
SIR 95%	0.010	0.010	0.005	0.526***	0.170	0.129	-0.264
#14 Stocks			(0.371)	(0.001)	(0.481)	(0.582)	(0.136)
SIR 90%	0.009	0.009	0.003	0.555***	0.284	-0.024	-0.406***
#28 Stocks			(0.461)	(0.000)	(0.166)	(0.902)	(0.007)
SIR 5%	0.025	0.025	0.020***	0.465***	0.093	0.180	-0.107
#14 Stocks			(0.000)	(0.002)	(0.694)	(0.432)	(0.535)
SIR 10%	0.021	0.021	0.016***	0.470***	0.108	0.123	-0.136
#28 Stocks			(0.000)	(0.000)	(0.487)	(0.414)	(0.229)
SIR 5%-SIR 95%	0.015	0.015	0.016**	-0.061	-0.077	0.051	0.157
			(0.013)	(0.729)	(0.788)	(0.855)	(0.455)
SIR 10%-SIR 90%	0.012	0.012	0.012***	-0.085	-0.176	0.147	0.270*
			(0.005)	(0.491)	(0.383)	(0.453)	(0.068)

Notes: Table 21 shows the regression analysis of monthly equal weighted highly and lightly shorted stock portfolios. Data source using Microsoft Excel from Thomson Reuters Canada Total Return Index DataStream Data. The regression equation run to obtain each portfolio in the table is as follows:

$$R_{port\,t} - R_{ft} = \alpha_{port} + \beta_{port}\left(R_m - R_f\right)_t + \gamma_{port}SMB_t + \delta_{port}HML_t + \eta_{port}MOM_t + \varepsilon_t \qquad (55)$$

The dependent variable is the equal weighted monthly portfolio excess return $R_{port\,t} - R_{ft}$ on highly or lightly shorted stock portfolios for the month subsequent to portfolio creation on stocks from the Thomson Reuters Canada Total Return Index. $\left(R_m - R_f\right)_t$ is the market risk premium, SMB_t is the excess return of a portfolio of small cap stocks over a portfolio of large cap stocks, HML_t is the excess return of a portfolio of high book to market stocks over low book to market books and MOM_t is the excess return of period t-1 winners' portfolio over t-1 losers' portfolio. The monthly data for the three Fama French factors and the one Momentum factor is from Kenneth French's Data Library. The portfolios SIR 95% and SIR 90% include stocks with short interest ratios from the 95th and 90th percentile respectively in month t-1. The portfolios SIR 5% and SIR 10% include stocks with short interest ratios from the 5th and 10th percentile respectively in month t-1. The portfolios are rebalanced each month based on new short interest data. Portfolio raw returns and excess returns over the risk-free rate are shown under the headings of Raw Return and Excess Return. The number of stocks in each portfolio can change each month and thus the average number of stocks in each portfolio is shown under the heading # stocks. I report p values in brackets under respective regression coefficients. Stars are used to indicate significant figures. * is significance at the 10% level, ** is significance at the 5% level and *** is significance at the 1% level. My sample period is based on monthly short interest data from February 2010 to July 2017. Data author's own calculation.

I next report the regression analysis results in **Table 21**. If we look at Columns (1) and (2) in Table 21, we do indeed see that the SIR 5% portfolio is the best performing portfolio with a raw and excess return of 2.5% per month. The raw and excess returns are virtually identical on a rounding basis due to the effectively zero bound interest rates. I again use the 3-month US treasury bill as the risk-free rate. The worst performing portfolio is the SIR 90% portfolio, which averages a return of 0.9% per month on both a raw and excess return basis. The performance of portfolios is relatively evenly balanced with the SIR 10% portfolio performing worse than the SIR 5% portfolio. On a raw and excess return basis the SIR 5% portfolio returns 2.5% a month, while the SIR 10% portfolio returns 2.1% a month. The SIR 95% and SIR 90% portfolio return 1% and 0.9% respectively a month, the SIR 95% slightly outperforms the SIR 90% portfolio, which shows that returns are not completely evenly distributed in the dataset.

We can see that for the case of the Canadian markets, Columns (1) and (2) show that lightly shorted stocks seem to outperform heavily shorted stocks, this is consistent with my findings for the American markets, as well as with Boehmer et al. (2010), Asquith et al. (2005) and Desai et al. (2002). It is generally accepted in the literature that lightly shorted stocks outperform heavily shorted stocks and my findings reinforce this statement. Compared to the American markets, the outperformance is less, so although lightly shorted stocks outperform heavily shorted stocks, the degree of outperformance will depend on the market. Overall as a whole the American market performed better than the Canadian market over the seven-year time period from February 2010 to July 2017.

Looking into the results of the regression we can first pay attention to the intercepts in Column (3). The intercepts denote the alpha of the Fama and French Three Factor Model with Momentum, how much the model has failed to specify on a risk adjusted basis. A positive alpha coefficient indicates the portfolio performed better than the model would have expected and a negative intercept coefficient indicates that the portfolio performed worse than the model would have expected. Fund managers will try to maximise the alpha of their portfolios, showing their clients that they are returning better than the level of risk they are taking. This is often used as a means to justify underperformance of a fund manager in relation to a market benchmark such as the S&P 500.

From Column (3) we can see that the lightly shorted portfolios hold a positive alpha which indicates they have all performed better than the model would have expected, however there is a larger alpha in the lightly shorted stock portfolios and a zero alpha in the heavily shorted stock portfolios. The best alpha is found in the SIR 5% portfolio of 2% and the worst alpha is found in the heavily shorted portfolios of 0%. In an ideal world, an investor would look to raw returns to judge which portfolio to hold, but some fund managers are adamant to put alpha above raw returns as seen in Del Guercio and Reuter (2014) given risk management is a central issue in portfolio choice. This phenomenon is often referred to as "fund managers chasing alpha" or "fund managers focusing on alpha". This can often be taken as an excuse for poor returns, where a fund manager can claim even though their return was lower than the benchmark (such as the S&P 500) they achieved a return higher than the level of risk they were taking, i.e. by holding a higher alpha compared to other fund managers. However, personally I believe it is more advisable to hold the portfolio with the highest raw return, as long as an investor's investment horizon is long term and an investor is not swayed by day to day market fluctuations which are necessary if the investor is accepting an equity risk premium.[76] If anything, the market fluctuations [77] provide great buying opportunities, as long as the fundamentals of the business remain sound and the investor's investment horizon is long term to recoup losses from investments made in overbought conditions.

Looking further into the regression at Column (4) we can see the betas of the individual portfolios from SIR 5% to SIR 10% are lower than the betas from the SIR 95% to SIR 90% portfolios, this is a consistent result.

76 The equity risk premium states that for receiving the average of 5% above investment grade bonds, you must be prepared to accept volatility. To ride out volatility is why this premium is justified for the investor.

77 In the short run the market is often driven by fear and greed, the idea of losing hard earnt money and the idea of missing out on potential gains, also known as FOMO (fear of missing out). This can cause securities to swing wildly between optimism and pessimism, the smart investor would use the fluctuations of the market to his advantage. Buying into pessimism and refraining from buying/selling into optimism. This bipolar nature of the market has been documented well by Graham (1949). The greatest losses have often been seen by buying poor businesses in times of good economic conditions (Graham, 1949).

If I was able to form larger portfolios I am sure more of the unsystematic risk of the portfolios would be removed resulting in a much smaller beta on the lightly shorted stocks and a much higher beta on the heavily shorted stocks. On a momentum basis Column (7) shows that the heavily shorted portfolio of SIR 90% exhibit much more negative momentum than the lightly shorted portfolios, this is consistent with the American markets and previous literature.

Regarding the long/short portfolios, a negative beta from Column (4) on the SIR 5%- SIR 95% portfolio is not as evident as in the US sample but is evident nonetheless. Looking at raw returns of the long/short portfolios it is still evident that a long only strategy is viable, since none of my lightly shorted portfolios produced a negative monthly raw return. The outperformance of lightly shorted stocks over heavily shorted stocks is very evident though. It is worth noting that it is difficult to compare long only portfolios with long/short portfolios, due to the short component borrowing costs. Even accounting for borrowing costs we can see that the long/short portfolios vastly underperform the long SIR 5% portfolio. Therefore going long the SIR 5% portfolio is advised.

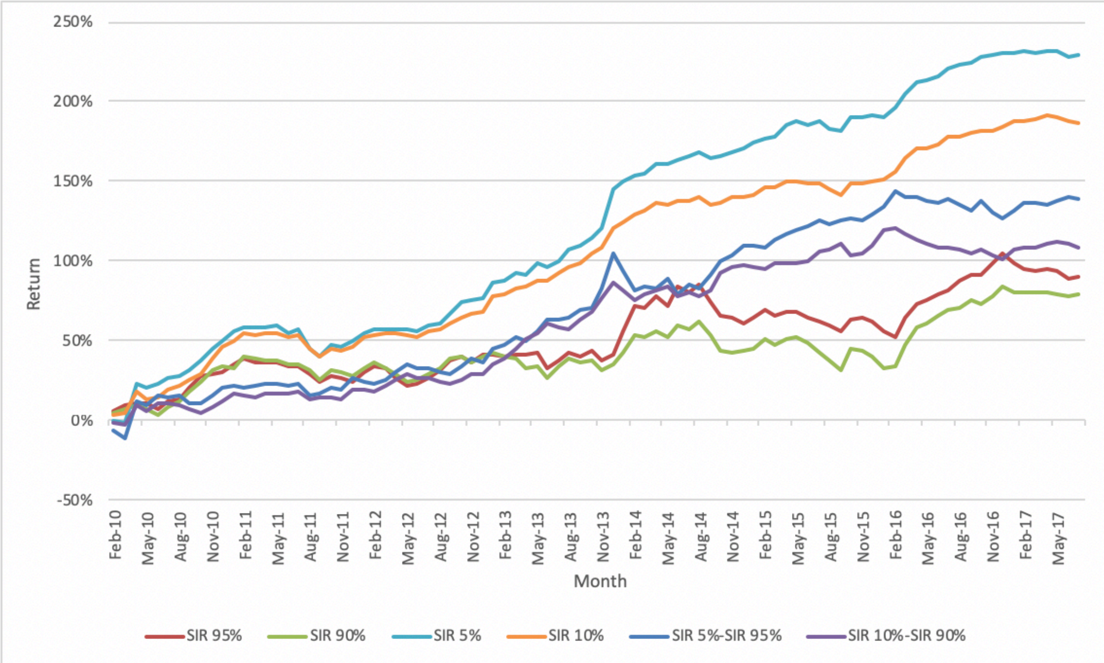
Table 22: Regression Statistics for Portfolios

	(1)	(2)	(3)	(4)
Portfolios	Multiple R	R Square	Adjusted R Square	Observations
SIR 95%	0.478	0.229	0.193	90
SIR 90%	0.591	0.349	0.319	90
SIR 5%	0.414	0.171	0.132	90
SIR 10%	0.573	0.328	0.297	90
SIR 5% - SIR 95%	0.106	0.011	-0.035	90
SIR 10% - SIR 90%	0.251	0.063	0.019	90

Notes: Table 22 shows the regression statistics for each portfolio. Again, we report for Multiple R, R Square, Adjusted R Square and Observations. Data author's own calculation.

Table 22 shows the regression statistics for portfolios. Looking at Table 22 in Column (4) there are 90 observations in each portfolio for the calendar months of February 2010 to July 2017 inclusive. We see in Column (2) a stronger R Square for my larger portfolios of SIR 90% and SIR 10%, as we imagined the smaller portfolios have more volatility and are going to produce a smaller R Square. The portfolio with the best R Square is the SIR 90% portfolio with a R Square of 0.349 and the portfolio with the worst R Square is the long/short SIR 5%-SIR 90% portfolio with a R Square of 0.011. This is in turn due to the low R Square of the SIR 5% portfolio.

Figure 21: Cumulative SIR Portfolio Returns for Canada



Notes: Data Source using Microsoft Excel from Thomson Reuters Canada Total Return Index DataStream Data. Figure 21 shows the cumulative returns of the different short interest ratio (SIR) portfolios over the period of the study.

Figure 21 shows the cumulative SIR portfolio returns for Canada. Looking at Figure 21 we can easily see that overtime the outperformance of the SIR 5% portfolio is evident and we can see that the SIR 5% portfolio outperforms the SIR 95% portfolio by a margin of over 100%. The SIR 5% performance outperforms the others, including the SIR 10% portfolio which has stocks that are included in the SIR 5% portfolio. The worst performing portfolio is the SIR 90% portfolio which for a period of two years (February 2010 to February 2012) posts a total return underperforming the S&P 500. Overall, we see that there is a clear performance premium of the SIR 5% and SIR 10% portfolios over the SIR 95% and SIR 90% portfolios. It is worth noting that in the case of Canada, the long/short portfolios outperform the heavily shorted portfolios by a small margin, again I do not recommend a shorting strategy based on the vast outperformance of the SIR 5% portfolio and SIR 10% portfolio over the other portfolios.

Table 23: Correlation Matrix for SIR Portfolio Returns for Canada

Portfolios	SIR 95%	SIR 90%	SIR 5%	SIR 10%	SIR 5%-SIR 95%	SIR 10%-SIR 90%
SIR 95%	1	0.886	0.411	0.491	-0.580	-0.630
SIR 90%	0.886	1	0.474	0.588	-0.416	-0.681
SIR 5%	0.411	0.474	1	0.892	0.504	0.232
SIR 10%	0.491	0.588	0.892	1	0.331	0.192
SIR 5%-SIR 95%	-0.580	-0.416	0.504	0.331	1	0.805
SIR 10%-SIR 90%	-0.630	-0.681	0.232	0.192	0.805	1

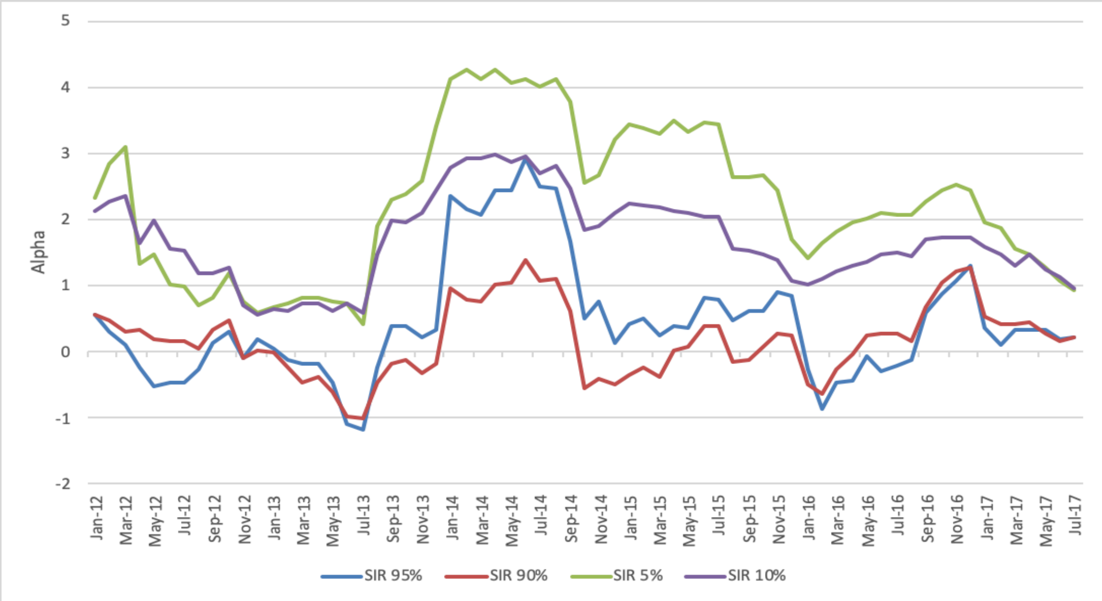
Notes: Table 23 shows the Correlation Matrix for SIR Portfolio Returns for Canada. This correlation matrix matches the return correlations between high, low and long/short SIR portfolios. A correlation of 1 indicates perfect positive correlation, a correlation of -1 indicates perfect negative correlation and a correlation of close to 0 indicates little to no correlation.

Table 23 shows the correlation matrix for SIR portfolio returns for Canada. The correlation matrix allows us to see how correlated returns are between different portfolios. We see the strongest positive correlation between the SIR 5% and SIR 10% portfolios of 0.892; the SIR 95% and SIR 90% portfolios also hold a strong positive correlation of 0.886. This is to be expected since the SIR 10% portfolio includes stocks from the SIR 5% portfolio and the SIR 90% portfolio includes stocks from the SIR 95% portfolio. Even with this knowledge, the correlations remain positively strong and show returns may very well be influenced from stocks in the extreme percentiles.

The strongest negative correlation of -0.681 is that of the SIR 90% portfolio and SIR 10%-SIR 90% portfolio, again formed due to the short component of the SIR 10%-SIR 90% portfolio. The SIR 10% long component of the SIR 10%-SIR 90% portfolio, means we have a correlation which is not perfectly negative.

The nature of volatility in equity markets meant they went through a great deal of changes over my sample period. Thus, in extension of my main results, I again attempt to explore if my findings are period specific or whether there are significant changes over time. To do this I conduct a 24-month rolling regression using the Fama and French Three Factor Model with Momentum shown in Equation (55) to report for both alpha coefficient (abnormal returns) and beta coefficient (volatility with respect to market) across my sample period. I conduct the rolling regression for SIR 95%, SIR 90%, SIR 5%, SIR 10%, SIR 5%-SIR 95% and SIR 10%-SIR 90% portfolios, this means I account for heavily shorted portfolios, lightly shorted portfolios and long/short portfolios.

Figure 22: SIR Portfolio Rolling Regression Alphas for Canada

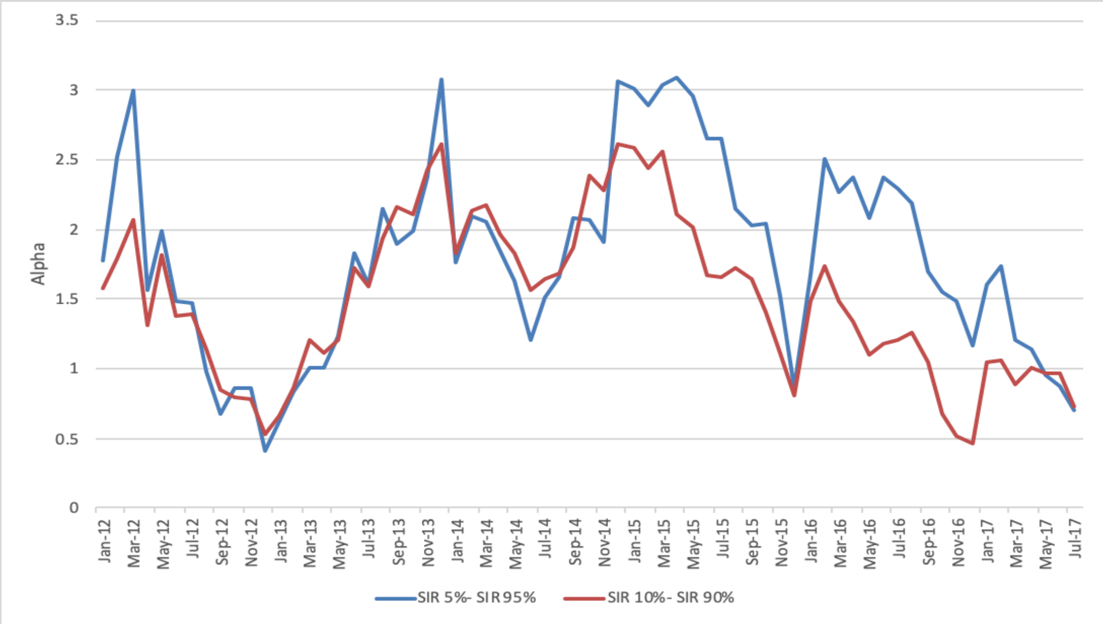


Notes: Figure 22 shows the SIR Portfolio Rolling Regression Alphas for Canada. The rolling regression is conducted across a rolling 24-month period from February 2010 to July 2017 using the Fama and French Three Factor Model with Momentum. The dates shown in Figure 22 account for the end period of that particular rolling regression, which is conducted on a monthly basis. I report for Alphas, which show the outperformance or underperformance of the portfolios with respect to the Fama and French Three Factor Model with Momentum benchmark.

Figure 22 shows the rolling regression alphas for Canada using the SIR 95%, SIR 90%, SIR 10% and SIR 5% portfolios. We see the outperformance of the lightly shorted portfolios over the heavily shorted portfolios in terms of alpha over my study period. There is a deal of correlation between alphas and portfolios, where lightly shorted portfolios follow heavily shorted portfolios in decreases in alpha and also increases in alpha. Again, this correlation could be down to risk-on and risk-off nature of equity markets, where money enters the equity market across both high and low short interest stocks and leaves the equity market again across both high and low short interest stocks.

The total range of alphas for the heavily shorted portfolios is between -1.1 and 3, the total range of alphas for the lightly shorted portfolios is between 0.5 and 4.2. The range of alphas are similar in both portfolio types, but the lightly shorted portfolios hold the higher mean and median alpha.

Figure 23: Long/Short SIR Portfolio Rolling Regression Alphas for Canada

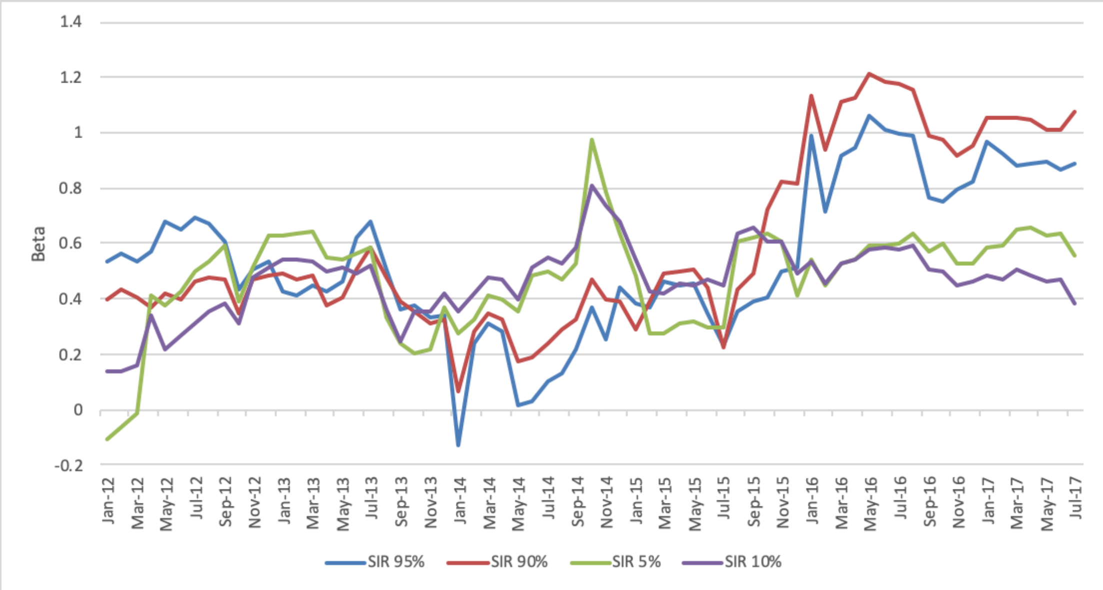


Notes: Figure 23 shows the Long/Short SIR Portfolio Rolling Regression Alphas for Canada. The rolling regression is conducted across a rolling 24-month period from February 2010 to July 2017 using the Fama and French Three Factor Model with Momentum. The dates shown in Figure 23 account for the end period of that particular rolling regression, which is conducted on a monthly basis. I report for Alphas, which show the outperformance or underperformance of the portfolios with respect to the Fama and French Three Factor Model with Momentum benchmark.

Figure 23 shows the rolling regression alphas for the Canada using the SIR 5%-SIR 95% and SIR 10%-SIR 90% portfolios. The long/short portfolio alphas account for the overperformance or underperformance of long/short portfolios with respect to the Fama and French Three Factor Model with Momentum benchmark.

The alphas for the long/short portfolios are much tighter than previous long/short portfolios of the US market in research questions 1 and 2. There are times when the SIR 5%-SIR 95% has the higher alpha and times when the SIR 10%-SIR 90% portfolio has the higher alpha. This suggests that the outperformance of the extreme lightly shorted stocks and the underperformance of the extreme heavily shorted stocks may not be as great. This could be due to the composition of stocks in the Canadian market as opposed to the US market, which there are fewer extreme differences in performance between heavily and lightly shorted stocks. There is a great deal of correlation between the long/short portfolios, which is to be expected as stocks in the SIR 5%- SIR 95% portfolio will be in the SIR 10%- SIR 90%.

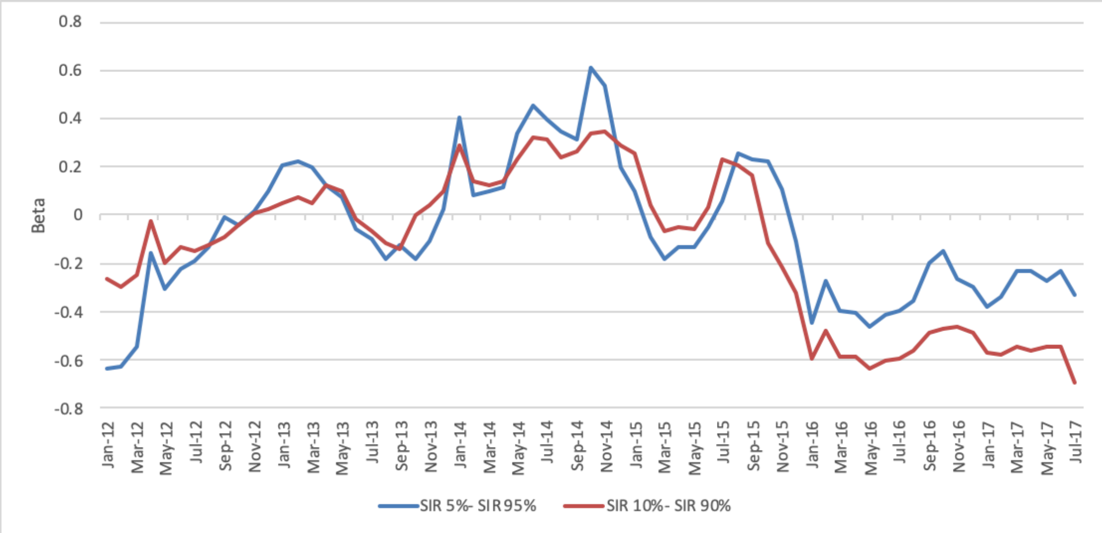
Figure 24: SIR Portfolio Rolling Regression Betas for Canada



Notes: Figure 24 shows the SIR Portfolio Rolling Regression Betas for Canada. The rolling regression is conducted across a rolling 24-month period from February 2010 to July 2017 using the Fama and French Three Factor Model with Momentum. The dates shown in Figure 24 account for the end period of that particular rolling regression, which is conducted on a monthly basis. I report for Betas, which show the volatility of the portfolios with respect the market.

Figure 24 shows the rolling regression betas for Canada using the SIR 95%, SIR 90%, SIR 10% and SIR 5% portfolios. Before May 2015 we see a good deal of correlation between portfolios, after this there is a divergence and the lightly shorted portfolios hold lower betas than the highly shorted portfolios. Throughout the study period the betas of all portfolios have been on a slight uptrend. The betas range from 1.2 to -0.1 for all portfolios, a beta over 1 is observed by the heavily shorted portfolios after May 2015, which is more volatile than the market portfolio. It is natural for the beta of these portfolios to vary, since the other factors of the Fama and French Three Factor Model with Momentum (such as SMB, HML or MOM) may be contributing more the performance of portfolios over the beta factor.

Figure 25: Long/Short SIR Portfolio Rolling Regression Betas for Canada



Notes: Figure 25 shows the Long/Short SIR Portfolio Rolling Regression Betas for Canada. The rolling regression is conducted across a rolling 24-month period from February 2010 to July 2017 using the Fama and French Three Factor Model with Momentum. The dates shown in Figure 25 account for the end period of that particular rolling regression, which is conducted on a monthly basis. I report for Betas, which show the volatility of the portfolios with respect the market.

Figure 25 shows the rolling regression betas for the Canada using the SIR 5%-SIR 95% and SIR 10%-SIR 90% portfolios. The long/short portfolio betas account for the volatility of long/short portfolios with respect to the market. We see a good deal of correlation between the two portfolios. The range of the betas of the portfolios is between 0.7 and -0.7. It is normal for the beta of these portfolios to be changing, since the other factors of the Fama and French Three Factor Model with Momentum could be influencing more the performance of portfolios over the beta factor.

My original research question was: Is the long/short strategy of Boehmer et al. (2010) still valid in an international OECD market that of Canada? I used Canada as a good proxy of an OECD country. From what we have observed, the original strategy of Boehmer et al. (2010) of going long the least shorted portfolio and going short the most shorted portfolio is not valid, since the SIR 95% has posted a positive raw return than the slight negative return the most shorted portfolio has posted in the study from Boehmer et al. (2010). The negative return was marginal at -0.1%, so going short based on that data is with caution as a flat return or slightly positive return would incur extra transaction fees while providing little or no benefit to the investor.

Based on the results of the Canadian Dataset it is advisable to go long the SIR 5% portfolio, where the average monthly raw return is at 2.5% and a risk adjusted alpha of 2%. This is the same strategy I advised in research question 2 for the American markets. What must be noted is that lightly shorted stocks do indeed outperform heavily shorted stocks, so if you hold onto lightly shorted stocks you will vastly outperform heavily shorted stocks and indeed the market return of the S&P 500.

Based on these reasons I believe that there is good news in short interest for the Canadian market if the correct strategy of going long lightly shorted portfolios and in particular going long the SIR 5% portfolio is followed. Again, I must mention the implementation of a strategy such as going long the SIR 5% portfolio is difficult for the average investor, as he or she must rebalance every month. In this case it is better since there are only 14 stocks in the portfolio, but the constant rebalancing of these 14 stocks can lead to transaction fees building up. If a fund manager is more concerned about the alpha of the portfolio,[78] the SIR 5% portfolio again also shows to hold better, though I believe as previously stated before raw return is more significant than alpha when judging which is the better portfolio to hold. However, many fund managers as seen in Del Guercio and Reuter (2014) chase Alpha to appease investors who demand risk-adjusted returns. Investors should want to hold the portfolio that ideally generates the highest return, rather than return on a risk adjusted basis.

My findings for the Canadian dataset are consistent when compared to the American dataset, the outperformance of lightly shorted stocks is evident. We have seen replicating a study outside of the US is challenging indeed due to data constraints of securities. This concludes my findings for research question 3.

6.4 Whether and to what extent short sales affect liquidity, price discovery, volatility and cross-section of stock returns? (Research Question 4)

This research question aimed to answer the following: Whether and to what extent short sales affect liquidity, price discovery, volatility and cross-section of stock returns? I look to explore whether the 2007-2009 financial crisis short sale ban leads to changes in stock volatility, liquidity and price discovery. I first report for portfolio descriptive statistics, then volatility results, then liquidity results and finally I report for price discovery results. The main results for volatility changes using a GARCH (1,1) model with a dummy variable can be seen in **Table 25**, the main results for liquidity changes using a Bid-Ask Spread model with a dummy variable can be seen in **Table 27** and the main results for the price discovery changes using the runs test can be seen in **Table 31**.

Table 24: Unbanned Portfolio and Banned Portfolio Descriptive Statistics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Portfolios	Mean	Standard Error	Median	Standard Deviation	Kurtosis	Skewness	Range	Maximum	Minimum
Unbanned Portfolio (Control)	0.000	0.001	0.000	0.022	4.843	0.458	0.194	0.114	-0.080
Banned Portfolio	-0.002	0.002	-0.004	0.034	4.404	0.883	0.304	0.187	-0.116

Notes: Table 24 shows the descriptive statistics for both the Unbanned and Banned Portfolios. Descriptive statistics show return on a daily basis for each portfolio. Total of 264 daily observations. Data author's own calculation.

78 Fund performance is often quoted with alpha values, the higher the past alpha of the fund, the better it has performed relative to risk taken. However, investors should be more concerned with the raw return they are achieving and in particular the fund manager fees they are paying. When a Vanguard S&P 500 Index fund has fees on average of 0.04% of capital invested per year, a fund manager needs to justify why an investor should be paying 2% in fees. These fees erode raw return in the long term, especially if returns are made from reinvesting dividends.

Table 24 shows the descriptive statistics for Unbanned and Banned Portfolios used in both volatility, liquidity and price discovery results. From the descriptive statistics in Table 24 we can see that both portfolios performed poorly during my study period. In Column (1) we see the mean portfolio return per day was either 0 or -0.2%, showing that my study occurred during a period of crisis and thus in a bear market. Stocks in general tend to rise in the long term, especially baskets of stocks where the risk is diversified across sectors. It is exceptional to see a sustained period of negative returns in stocks, showing how deep the financial crisis truly was. In Column (2) we see the standard error for both portfolios remains relatively small. In Column (3) we see the median for the Unbanned Portfolio is very similar to the mean for the Unbanned Portfolio showing a lack of skewness is the distribution of the returns. The Banned Portfolio has a lower median, emphasising a higher skew than the Unbanned Portfolio. In Column (4) we see the Banned Portfolio has the higher standard deviation showing that it is more volatile than the Unbanned Portfolio, which is not surprising since financial stocks were the worst hit during the crisis.

The kurtosis shown in Column (5) is marginally higher in the Unbanned Portfolio and is an indication of fat tails, though both kurtosis results do not show a large degree of fat tails in the sample. We see in Column (6) that the Banned Portfolio exhibits more negative skew than the Unbanned Portfolio. The range of the Banned Portfolio shown in Column (7) is much higher, indicative of the higher standard deviation in the sample. In Columns (8) and (9) the minimum and maximum returns are higher in the Banned Portfolio, again shown by the range and standard deviation of returns.

So, in general the Unbanned Portfolio has outperformed the Banned Portfolio and exhibited less volatility than the Banned Portfolio. As an investor it has been better to be not invested in financial stocks during the 2007-2009 financial crisis, though many would have not known that the housing market in the US was overheating. However, there are speculators who are willing to take on catastrophic risk and could have gone well investing in financial stocks at the bottom of the bear market in March 2009.

6.4.1 Volatility Results

Table 25: Unbanned Portfolio and Banned Portfolio GARCH (1,1) Model with Dummy Variable

		Unbanned Portfolio (Control)		Banned Portfolio	
		(1)	(2)	(3)	(4)
Equation	Variable	Coefficient	P-value	Coefficient	P-value
Mean Equation	Constant	0.000	(0.522)	0.000	(0.921)
	FTSE 100 daily return	0.730***	(0.000)	1.182***	(0.000)
Conditional Equation	Constant	0.000**	(0.014)	0.000	(0.226)
	ARCH 1	0.237***	(0.000)	0.092***	(0.005)
	GARCH 1	0.776***	(0.000)	0.877***	(0.000)
	Dummy	0.002	(0.180)	-0.001	(0.737)

Notes: Table 25 shows the GARCH (1,1) model with dummy variable results for both Unbanned and Banned Portfolios. I report for both Mean and Conditional Equations. With coefficients I report for the Constant coefficient, FTSE 100 daily return coefficient, ARCH 1 coefficient, GARCH 1 coefficient and Dummy coefficient. P-values are given in brackets and indicate significance levels. I use a star system like previous to denote significance, *** is significance at the 1% level, ** is significance at the 5% level and * is significance at the 10% level. I investigate banned stocks but use unbanned stocks as a control variable. Data author's own calculation, calculated using STATA 15. The following model is used for the production of these results:

$$R_t = \alpha_t + \beta_i R_{mt} + \varepsilon_t \tag{56}$$

$$\varepsilon_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + D_{SSB} \tag{57}$$

The model above represents an GARCH (1,1) model with a dummy variable to take note of changes in volatility between the no short sale ban and short sale ban time periods. R_t is the portfolio return at time t and R_{mt} is the FTSE 100 daily return at time t. ε_{t-1}^2 is the ARCH effect at time t-1 and ε_t^2 is the ARCH effect at time t. σ_{t-1}^2 is the GARCH effect at time t-1. D_{SSB} is the dummy variable. ε_t is the error term. I take daily observations between January 3rd 2008 and January 16th 2009, total of 264 daily observations.

I report for volatility results in **Table 25** next. It is difficult to control the industry effect around the ban, since financial stocks were subject to a short sale ban. This is worth considering when evaluating the volatility results that the Unbanned Portfolio (control) may have trouble in being a true control. However, this doesn't take away the fact that the results from my Banned Portfolio can be evaluated and compared to the best control that we do have.

I find a very good fit for both GARCH models in Table 25, with both ARCH 1, GARCH 1 and FTSE 100 daily return terms holding statistical significance at the 1% level. The constant term of the conditional equation shown in Column (1) holds significance at the 5% level. We can see that both portfolios can be modelled using GARCH models, though the Unbanned Portfolio can fit the EGARCH specification[79] and this is evident from the volatility clustering and asymmetric nature of returns shown in the unbanned portfolio.

[79] Appendix 1 shows the replication of Table 25 using the EGARCH specification. Using AIC and BIC, the best fitting EGARCH model is EGARCH (2,3). This EGARCH specification is built with 2 ARCH lags and 3 GARCH lags.

The main term used to measure change in the dummy variable in both portfolios. A change in the dummy variable with a level of significance at the 10% level means we can be confident of increases or decreases in volatility between the two time periods. In my case both banned and unbanned portfolios are outside this significance level (shown in Columns (2) and (4)), therefore from my study we have seen that the ban of short selling in financial stocks has little effect on volatility. This leads me, like many other studies, to question the effectiveness of the short sale ban and whether regulatory authorities should be implementing a ban such as this in the first place. Many other studies find that the short sale ban either leads to slight increases in volatility or no effect on volatility at all.

What is also worth noting is the dummy variable of the unbanned stocks in Column (1) also did not have a statistically significant change in volatility, adding to my conclusion that the short sale ban may not be effective at all. This however has not stopped governments implementing short sale bans and this study along with many others such as Ho (1996), Chang et al. (2007) and Lobanova et al. (2010) question their significance. Many other studies have also shown short sale bans to lead to volatility increases, further adding to their ineffectiveness.

Both GARCH models fail to specify a constant term for the Mean Equation, however the constant term for the Conditional Equation in the Unbanned Portfolio is significant. Other than that, the other constant terms are taken to be zero. The FTSE 100 daily return matches my portfolios to the market return, we see my banned stock portfolio holding a lesser match to the return of the FTSE 100 in general. The asymmetric effects are more significant on the unbanned stock portfolio over the banned stock portfolio.

It is worth looking at why there was not a statistical change in volatility, at any significant level, since we were going through a period of crisis which deepened once the short sale ban was in place, further volatility may have resulted from this crisis to offset the downward selling pressure of short sellers. Since short sale bans usually take place during crisis's, this behaviour is likely to persist. If a short sale ban were to be implemented in say a non-crisis period, volatility effects may be present, however there has not been a case in financial history where a short sale ban has been put in place without negative downward pressure already present in the market. The relationship of volatility with short sale bans in crisis and non-crisis periods is however of note and in future study is worth investigating.

Table 26: Asset Management Portfolio, Insurance Portfolio and Banking Portfolio GARCH (1,1) Model with Dummy Variable

		Asset Management Portfolio		Insurance Portfolio		Banking Portfolio	
		(1)	(2)	(3)	(4)	(5)	(6)
Equation	Variable	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
Mean Equation	Constant	0.001	(0.731)	0.001	(0.753)	0.000	(0.978)
	FTSE 100 daily return	0.893***	(0.000)	1.154***	(0.000)	1.469***	(0.000)
Conditional Equation	Constant	0.000	(0.289)	0.000	(0.157)	0.000***	(0.009)
	ARCH 1	0.138***	(0.004)	0.112***	(0.007)	0.327***	(0.010)
	GARCH 1	0.844***	(0.000)	0.874***	(0.000)	0.456***	(0.005)
	Dummy	0.001	(0.738)	0.001	(0.823)	-0.003	(0.395)

Notes: Table 26 shows the GARCH (1,1) model with dummy variable results for Asset Management, Insurance and Banking Portfolios. Portfolios are created from the main revenue stream of either asset management, insurance or banking using stocks in the Banned Portfolio. I report for both Mean and Conditional Equations. With coefficients I report for the Constant coefficient, FTSE 100 daily return coefficient, ARCH 1 coefficient, GARCH 1 coefficient and Dummy coefficient. P-values are given in brackets and indicate significance levels. I use a star system like previous to denote significance, *** is significance at the 1% level, ** is significance at the 5% level and * is significance at the 10% level. Data author's own calculation, calculated using STATA 15. The following model is used for the production of these results:

$$R_t = \alpha_t + \beta_i R_{mt} + \varepsilon_t \tag{58}$$

$$\varepsilon_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + D_{SSB} \tag{59}$$

The model above represents an GARCH (1,1) model with a dummy variable to take note of changes in volatility between the no short sale ban and short sale ban time periods. R_t is the portfolio return at time t and R_{mt} is the FTSE 100 daily return at time t. ε_{t-1}^2 is the ARCH effect at time t-1 and ε_t^2 is the ARCH effect at time t. σ_{t-1}^2 is the GARCH effect at time t-1. D_{SSB} is the dummy variable. ε_t is the error term. I take daily observations between January 3rd 2008 and January 16th 2009, total of 264 daily observations.

Table 26 extends the main results further by splitting the Banned Portfolio into 3 distinctive portfolios and replicating the results of Table 25. This allows us further analysis into the behaviour of the Banned Portfolio. The Banned Portfolio consists of financial stocks and I split the stocks in the portfolio into 3 categories of Asset Management, Insurance and Banking depending on the stock's main line of business and revenue stream.

The Asset Management Portfolio consists of the firms St James's Place and Standard Life Aberdeen. These firms focus their core business on asset management of pensions and private investors. The Insurance Portfolio consists of the firms RSA Insurance Group, Prudential, Legal and General, Aviva and Admiral Group. These firms focus their core business on life insurance, car insurance and reinsurance. The Banking Portfolio consists of the firms Royal Bank of Scotland Group, HSBC Holdings, Standard Chartered, Barclays and Lloyds Banking Group. These firms focus their core business on retail banking and investment banking. All of these stocks are components of the Banned Portfolio.

We again see a good fit for all portfolios in Table 26, the GARCH specification holds well with the Banned Portfolio and as expected it holds well with the sub portfolios created from it. All GARCH and ARCH terms are within the 1% significance level, showing a great deal of volatility clustering in the data.

All portfolios have a lack of significance with the Dummy term, showing that all sectors have not had changes in volatility, between the non-ban period and the short sale ban period. The Asset Management Portfolio best models the FTSE 100 daily return, where the Banking Portfolio least models the FTSE daily return.

ARCH effects are more prominent in the Banking Portfolio, while GARCH effects are more prominent in the Asset Management Portfolio and Insurance Portfolio. The Asset Management and Insurance Portfolio are similar in behaviour, while the Banking Portfolio is much more dependent on the GARCH process over the ARCH process driving the returns.

6.4.2 Liquidity Results

Table 27: Unbanned Portfolio and Banned Portfolio Bid-Ask Spread Model with Dummy Variable

	Unbanned Portfolio (Control)		Banned Portfolio	
	(1)	(2)	(3)	(4)
Variable	Coefficient	P-value	Coefficient	P-value
Constant	1.784***	(0.000)	1.782***	(0.000)
Daily Average Return Squared	0.815***	(0.001)	0.093	(0.580)
Daily Average Trading Volume	0.000***	(0.000)	0.000***	(0.000)
Daily Average Excess Trading Volume	0.000	(0.912)	0.000	(0.210)
Dummy	0.008***	(0.000)	0.019***	(0.000)

Notes: Table 27 shows the Bid-Ask Spread model with dummy variable results for both Unbanned and Banned Portfolios. I report for the Constant coefficient, Daily Average Return Squared coefficient, Daily Average Trading Volume coefficient, Daily Average Excess Trading Volume and Dummy coefficient. P-values are given in brackets and indicate significance levels. I use a star system on coefficients like previous to denote significance, *** is significance at the 1% level, ** is significance at the 5% level and * is significance at the 10% level. I investigate banned stocks but use unbanned stocks as a control variable. Data author's own calculation, calculated using STATA 15. The following model is used for the production of these results:

$$S_t = c + \beta_0 \alpha r_t^2 + \beta_1 \alpha v_{it} + \beta_2 e v_{it} + \beta_3 D_{SSB} + \varepsilon_t \qquad (60)$$

The model above represents a Bid-Ask Spread Model with a dummy variable to take note of changes in liquidity between the no short sale ban and short sale ban time periods. S_t is the average spread for Unbanned or Banned Portfolio at time t, c is the constant, αr_t^2 is the daily average return of the Unbanned or Banned Portfolio squared at time t, αv_{it} is the daily average volume for the Unbanned or Banned Portfolio at time t, ev_{it} is the excess daily average volume of the Unbanned or Banned Portfolio at time t. D_{SSB} is the dummy variable. ε_t is the error term. I take daily observations between January 3rd 2008 and January 16th 2009, total of 264 daily observations.

I next report my liquidity results. It is difficult to control the industry effect around the ban, since financial stocks were subject to a short sale ban. This is worth considering when evaluating the liquidity results that the Unbanned Portfolio (control) may have trouble in being a true control. However, this doesn't take away the fact that the results from my Banned Portfolio can be evaluated and compared to the best control that we do have.

The main results for liquidity changes using a Bid-Ask Spread Model with a dummy variable can be seen in **Table 27**. I report for both Banned and Unbanned Portfolios in regard to liquidity.

Looking first in Column (1) we see a dummy coefficient of 0.008 and a significant P-Value at the 1% significance level. This indicates that the Unbanned Portfolio has a slight deterioration in liquidity between the ban and the unban periods. What this tells us is that a 1% increase in the daily spread of the 12 unbanned stocks leads to a 0.8% increase in the underlying liquidity of these stocks.

Next looking at Column (3) we see a dummy coefficient of 0.019 and a significant P-Value at the 1% significance level. This indicates that the Banned Portfolio has a greater deterioration in liquidity between the ban and the unban periods compared to the Unbanned Portfolio. What this tells us is that a 1% increase in the daily spread of the 12 banned stocks leads to a 1.9% increase in the underlying liquidity of these stocks. The short sale ban has affected the Banned Portfolio more than the Unbanned Portfolio, thus from this study we can see that the short sale ban has led to a deterioration in liquidity while accounting for other variables.

This result is consistent with other studies such as Zhisheng et al. (2018) and Alves et al. (2016), where liquidity has been deteriorated by imposing short sale bans. As market participants are removed from the market, bid-ask spreads widen and this leads to a loss in liquidity. Central banks have been adamant to impose short sale bans in periods of market turmoil, but my study like many others questions whether this policy has real intended impact rather than leading to a deterioration in market efficiency. I, along with many others, advise that trading bans would be more effective than short sale bans in order to stabilise markets.

Table 28: Unbanned Portfolio and Banned Portfolio Regression Statistics for Bid-Ask Spread Model

	(1)	(2)	(3)	(4)
Portfolios	Multiple R	R Square	Adjusted R Square	Observations
Unbanned Portfolio (Control)	0.815	0.664	0.659	264
Banned Portfolio	0.816	0.665	0.660	264

Notes: Table 28 shows the regression statistics for each portfolio for the Bid-Ask Spread Model. We report for Multiple R, R Square, Adjusted R Square and Observations. Data author's own calculation.

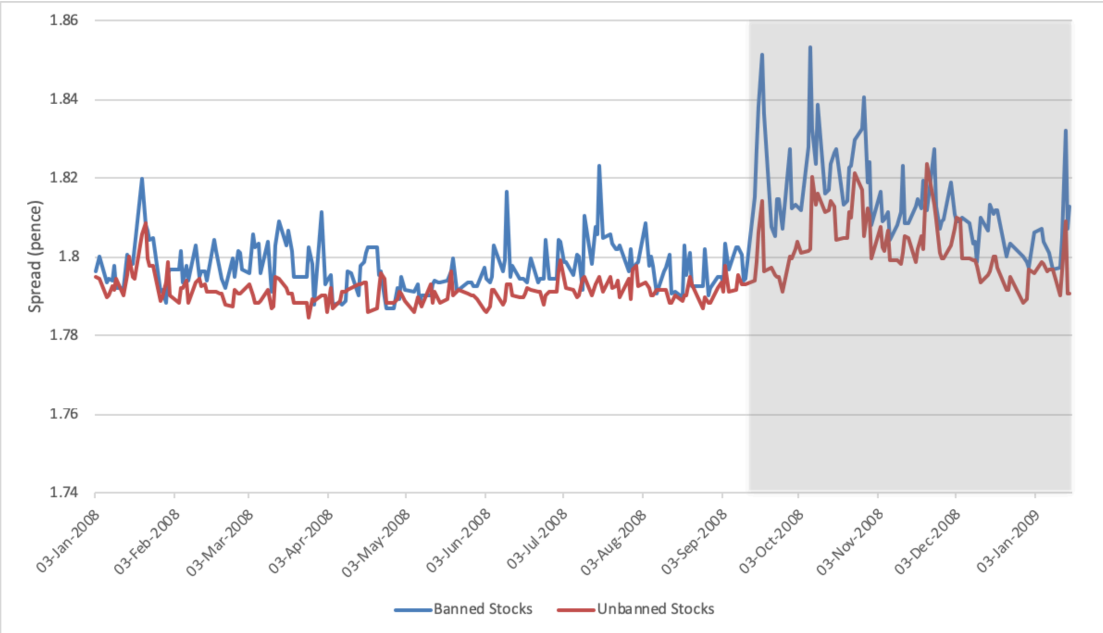
Table 28 shows the regression statistics for both Banned and Unbanned Portfolios for the Bid-Ask Spread Model. If we look at Column (2) we see the value of R Square which indicates how well the bid-ask spread model explains the regression data. The R Square for the Unbanned Portfolio is 0.664 and the R Square for the Banned Portfolio is 0.665, both levels of R Square indicate a reasonably good fit for the Bid-Ask Spread Model given that R Square is quoted between 0 and 1, where 0 is no fit and 1 is exceptional fit. We can be confident that the factors in the Bid-Ask Spread Model of return, trading volume and excess trading volume are contributing to the spread of securities in the portfolios. It is worth noting that the R Square values for both portfolios are very similar, again highlighting the point that these factors of return, trading volume and excess trading volume are contributing to the spread.

Looking at Column (3) we see the Adjusted R Square, which is a modified version of R Square that has been adjusted for the number of predictors in the model. The Adjusted R Square will increase if a new term will improve the model more than would be expected by chance alone.

We see the Unbanned Portfolio has an Adjusted R Square of 0.659 and the Banned Portfolio has an Adjusted R Square of 0.660. Both figures are similar to their R Squares respectively, which show the significance of terms affecting the model.

If we look at Column (4) we see the number of observations in the regression, which is 264 for both portfolios. As the number of observations increase, so does the predictive nature of a model, in an ideal world we would have an unlimited number of observations to better the fit of a model. The higher we can have the observations, the more reliable the fit of a model will be when forecasting.

Figure 26: Bid-Ask Spread using Corwin and Schultz (2012) Estimator for Unbanned and Banned Portfolios



Notes: Figure 26 shows the bid-ask spread between Unbanned and Banned Stock Portfolios. Spread is calculated using the Corwin and Schultz (2012) Estimator. Spread is given in pence and is the difference between the bid and the ask price. Spread is calculated using the following formula:

$$S = \frac{2(e^{\alpha}-1)}{(1+e^{\alpha})} \quad \text{Where: } \alpha = \frac{\sqrt{2\beta}-\sqrt{\beta}}{3-2\sqrt{2}} - \sqrt{\frac{\gamma}{3-2\sqrt{2}}}, \beta = E\left\{\sum_{j=0}^{1}\left(ln\frac{H_{t+j}^{0}}{L_{t+j}^{0}}\right)^2\right\}, \gamma = E\left\{\sum_{j=0}^{1}\left(ln\frac{H_{t,t+1}^{0}}{L_{t,t+1}^{0}}\right)^2\right\} \tag{61}$$

Given H_t is high price at time t, L_t is low price at time t, $H_{t,t+1}$ is highest price between time t and time t+1, $L_{t,t+1}$ is lowest price between time t and time t+1. High low-price parameters are used to calculate alpha, beta and gamma coefficients. Spread denoted by S is calculated using alpha, beta and gamma coefficients respectively on a daily basis for 264 observations. Area in grey indicates period of short sale ban. Data author's own calculation.

Figure 26 shows the bid-ask spread for both Banned and Unbanned Portfolios that has been taken using the Corwin and Schultz (2012) bid-ask spread estimator. We see that in general the Unbanned Portfolio has held the lower spread of the two portfolios. We also see that once the short sale ban takes place, both portfolios experience increases in spread but the Banned Portfolio experiences the much larger increase in spread. The short sale ban period is indicated with the grey area shown in Figure 26. This large increase in spread for the Banned Portfolio is attributed to the lack of short sellers in the market. It is clearly evident from Figure 26 that the short sell ban has affected spreads for the both the Banned and the Unbanned Portfolios.

Spreads are quoted in pence sterling and represent the difference between the bid and the ask price on average for each portfolio on a given day. A larger spread is beneficial for investors over day traders, as investors have a much longer holding period than day traders. A large spread may push many day traders out of securities, lowering volumes in the market.

Spreads for FTSE 100 stocks are generally small[80], due to the large amount of volumes entering the market. We see that the average spread is 1.8 pence for the Banned Portfolio and 1.79 pence for the Unbanned Portfolio across the entire 264 days of my study. During the short sale ban period, the average spread for the Banned Portfolio rises to 1.81 pence and the average spread for the Unbanned Portfolio rises to 1.8 pence. The Banned Portfolio experiences a much larger rise in spreads compared to the Unbanned Portfolio.

Table 29: Asset Management Portfolio, Insurance Portfolio and Banking Portfolio Bid-Ask Spread Model with Dummy Variable

	Asset Management Portfolio		Insurance Portfolio		Banking Portfolio	
	(1)	(2)	(3)	(4)	(5)	(6)
Variable	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
Constant	1.787***	(0.000)	1.786***	(0.000)	1.793***	(0.000)
Daily Average Return Squared	1.880***	(0.000)	0.362	(0.202)	-0.457*	(0.053)
Daily Average Trading Volume	0.000*	(0.083)	0.000	(0.130)	0.000	(0.686)
Daily Average Excess Trading Volume	0.000	(0.661)	0.000	(0.969)	0.000***	(0.002)
Dummy	0.013***	(0.000)	0.018***	(0.000)	0.016***	(0.000)

Notes: Table 29 shows the Bid-Ask Spread Model with dummy variable results for Asset Management, Insurance and Banking Portfolios. Portfolios are created from the main revenue stream of either asset management, insurance or banking using stocks in the Banned Portfolio. I report for the Constant coefficient, Daily Average Return Squared coefficient, Daily Average Trading Volume coefficient, Daily Average Excess Trading Volume and Dummy coefficient. P-values are given in brackets and indicate significance levels. I use a star system on coefficients like previous to denote significance, *** is significance at the 1% level, ** is significance at the 5% level and * is significance at the 10% level. Data author's own calculation, calculated using STATA 15. The following model is used for the production of these results:

$$S_t = c + \beta_0 \alpha r_t^2 + \beta_1 \alpha v_{it} + \beta_2 e v_{it} + \beta_3 D_{SSB} + \varepsilon_t \quad (62)$$

The model above represents a Bid-Ask Spread Model with a dummy variable to take note of changes in liquidity between the no short sale ban and short sale ban time periods. S_t is the average spread for the Asset Management, Insurance or Banking Portfolio at time t, c is the constant, αr_t^2 is the daily average return of the Asset Management, Insurance or Banking Portfolio squared at time t, αv_{it} is the daily average volume for the Asset Management, Insurance or Banking Portfolio at time t, ev_{it} is the excess daily average volume of the Asset Management, Insurance or Banking Portfolios at time t. D_{SSB} is the dummy variable. ε_t is the error term. I take daily observations between January 3rd 2008 and January 16th 2009, total of 264 daily observations.

80 I would expect spreads to be wider for small cap stocks, I would also expect a short sale ban to affect small cap stocks more negatively than large cap stocks, due to the lower levels of volumes in small cap stocks. The FTSE 100 holds the most liquid stocks on the London Stock Exchange, one of the largest stock market exchanges in the world by market capitalisation.

Table 29 extends the main results further by splitting the Banned Portfolio into 3 distinctive portfolios and replicating the results of Table 27. This allows us further analysis into the behaviour of the Banned Portfolio. The Banned Portfolio consists of financial stocks and I split the stocks in the portfolio into 3 categories of Asset Management, Insurance and Banking depending on the stock's main line of business and revenue stream.

The Asset Management Portfolio consists of the firms St James's Place and Standard Life Aberdeen. The Insurance Portfolio consists of the firms RSA Insurance Group, Prudential, Legal and General, Aviva and Admiral Group. The Banking Portfolio consists of the firms Royal Bank of Scotland Group, HSBC Holdings, Standard Chartered, Barclays and Lloyds Banking Group.

In terms of the Dummy Variable coefficient, shown in Columns (1), (3) and (5), all 3 portfolios have statistically significant changes in liquidity at the 1% level. The greatest change in liquidity is exhibited by the Insurance Portfolio and the least change in liquidity is exhibited by the Asset Management Portfolio.

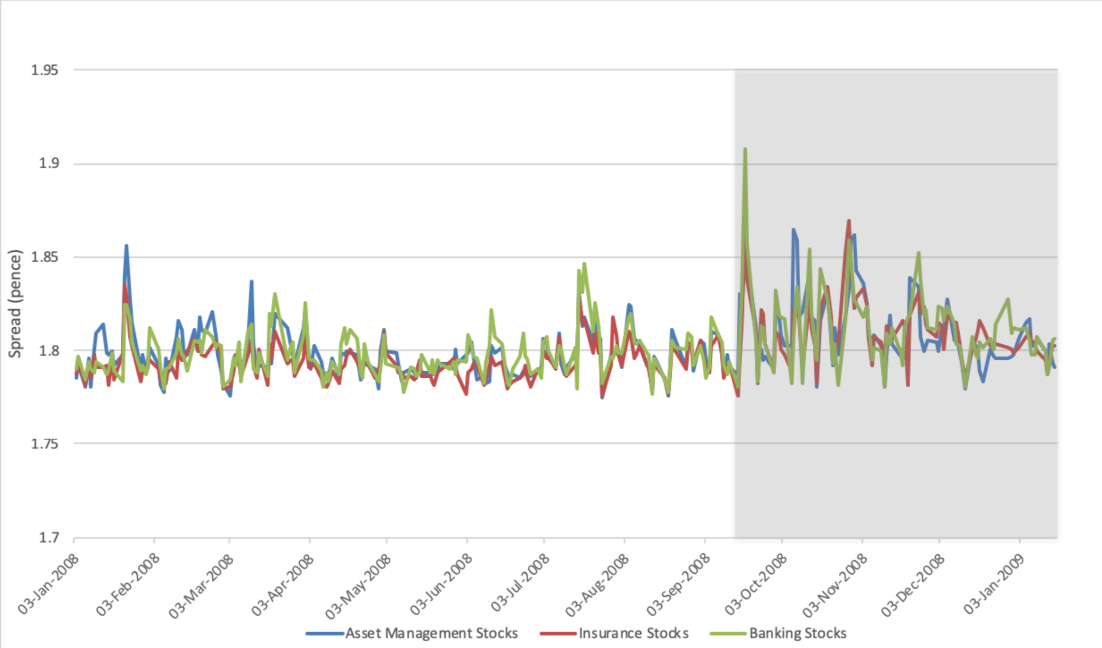
Table 30: Asset Management Portfolio, Insurance Portfolio and Banking Portfolio Regression Statistics for Bid-Ask Spread Model

	(1)	(2)	(3)	(4)
Portfolios	Multiple R	R Square	Adjusted R Square	Observations
Asset Management Portfolio	0.527	0.278	0.267	264
Insurance Portfolio	0.554	0.307	0.296	264
Banking Portfolio	0.594	0.352	0.343	264

Notes: Table 30 shows the regression statistics for Asset Management, Insurance and Banking Portfolios for the Bid-Ask Spread Model. We report for Multiple R, R Square, Adjusted R Square and Observations. Data author's own calculation.

Table 30 shows the regression statistics for the Asset Management, Insurance and Banking Portfolios. Looking at Table 30, we see the that the highest R Square shown in Column (2) is held by the Banking Portfolio of 0.352 and the lowest R Square (again, shown in Column (2)) is held by the Asset Management Portfolio. The observations remain the same as we are splitting one of my previous portfolios into 3 distinct portfolios based on firm revenue type. It is interesting to note that the Banking Portfolio fits my liquidity model the best.

Figure 27: Bid-Ask Spread using Corwin and Schultz (2012) Estimator for Asset Management Insurance and Banking Portfolios



Notes: Figure 27 shows the Bid-Ask Spread between Asset Management, Insurance and Banking Portfolios. Spread is calculated using the Corwin and Schultz (2012) Estimator. Spread is given in pence and is the difference between the bid and the ask price. Spread is calculated using the following formula:

$$S = \frac{2(e^{\alpha}-1)}{(1+e^{\alpha})} \quad \text{Where: } \alpha = \frac{\sqrt{2\beta}-\sqrt{\beta}}{3-2\sqrt{2}} - \sqrt{\frac{\gamma}{3-2\sqrt{2}}}\,, \beta = E\left\{\sum_{j=0}^{1}\left(ln\frac{H_{t+j}^{0}}{L_{t+j}^{0}}\right)^2\right\}, \gamma = E\left\{\sum_{j=0}^{1}\left(ln\frac{H_{t,t+1}^{0}}{L_{t,t+1}^{0}}\right)^2\right\} \qquad (63)$$

Given H_t is high price at time t, L_t is low price at time t, $H_{t,\,t+1}$ is highest price between time t and time t+1, $L_{t,\,t+1}$ is lowest price between time t and time t+1. High low-price parameters are used to calculate alpha, beta and gamma coefficients. Spread denoted by S is calculated using alpha, beta and gamma coefficients respectively on a daily basis for 264 observations. Area in grey indicates period of short sale ban. Data author's own calculation.

Figure 27 shows the bid-ask spread using Corwin and Schultz (2012) Estimator for Asset Management, Insurance and Banking Portfolios. Looking at the spreads in Figure 27, we see the three portfolios have relatively similar spreads. However, there are times when the Banking Portfolio has spreads which widen significantly compared to the Asset Management Portfolio and Insurance Portfolio. It is evident that during the short sale ban period, the spreads of all 3 portfolios significantly widen, contributing to the larger spreads seen in the Banned Portfolio over the Unbanned Portfolio. Overall it seems the Banking Portfolio has generally the widest spreads across my study period, followed by the Insurance Portfolio and finally the Asset Management Portfolio. It is also seen that stocks generally in periods of exceptional market turmoil experience wider spreads as noted in Clifton and Snape (2008). This is emphasised further in the Banking Portfolio as the crisis is specific to those stocks carrying housing debt loans.

6.4.3 Price Discovery Results

I first report for the runs in each portfolio across my study period of 3rd January 2008 to 16th January 2009. A run constitutes a consecutive succession of returns that are of the same sign. A zero return in a day continues the run and a run only ends when the sign of the return changes, therefore the smallest run can be of length 1 in absolute terms and the longest run can be of infinite length in absolute terms. Runs are limited to integer intervals, so it is not possible to have half a run or a quarter of a run. A zero run is also not defined, as zero is not positive or negative.

It is difficult to control the industry effect around the ban, since financial stocks were subject to a short sale ban. This is worth considering when evaluating the price discovery results that the Unbanned Portfolio (control) may have trouble in being a true control. However, this doesn't take away the fact that the results from my Banned Portfolio can be evaluated and compared to the best control that we do have.

Table 31 shows the main results of the Unbanned Portfolio (control) and Banned Portfolio run length in the non-ban period and short sale ban period. **Figure 28** shows a graphical representation of the Banned Portfolio and Unbanned Portfolio runs distribution for the non-ban period. **Figure 29** shows a graphical representation of the Banned Portfolio and Unbanned Portfolio runs distribution for the short sale ban period. Both Figure 28 and Figure 29 are graphical representations of the results in Table 31.

Table 31: Unbanned Portfolio and Banned Portfolio Run Length in Non-Ban Period and Short Sale Ban Period

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Period	Portfolio	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6
Non-Ban Period	Unbanned Portfolio (Control)	0	0	1	0	0	3	6	12	32	34	13	4	0	2	1
	Banned Portfolio	0	0	0	0	4	5	6	11	21	23	16	4	3	0	0
Short Sale Ban Period	Unbanned Portfolio (Control)	0	0	0	0	0	2	3	9	6	9	7	3	1	0	1
	Banned Portfolio	1	0	0	0	1	2	2	3	11	11	3	3	1	0	1

Notes: Table 31 shows the runs for the Unbanned Portfolio and Banned Portfolio for both Non-Ban Period and Ban Period for the 2008 Financial Crisis Short Sale Ban in the United Kingdom. Non-ban period includes daily returns from 2nd January 2008 to 18th September 2008 and the short sale ban period includes daily returns from 19th September 2008 to 16th January 2009. A run constitutes a consecutive sequence of either positive or negative daily returns for each portfolio. A zero-percentage daily return constitutes the run carrying on until there is a sign change.

Looking at **Table 31**, we initially see that the distribution of runs across the time period is wide with the maximum negative run length recorded at -9 and the maximum positive run length recorded at 6. This shows we are in a period of volatility, where wild downswings are often followed by wild upswings. This sort of market is beneficial in particular for speculators over investors, who can take advantage of these swings.

During the non-ban period, we see that both portfolios exhibit a normal distribution in runs (which is to be expected). Looking at Columns (9) and (10), we see the vast majority of the runs for both portfolios in the non-ban period are of run length 1 and -1.

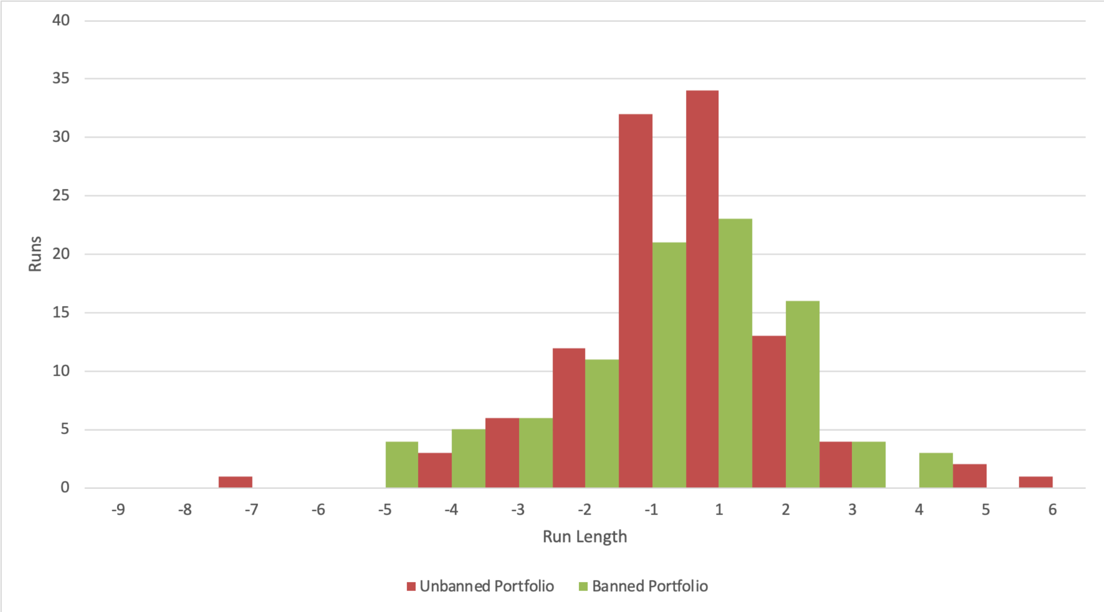
I consider runs of length 1 to 4 to be normal and runs of length 5 and above to be in the tail of the distribution, i.e. a fluctuation that is big and not considered normal, this is confirmed with a percentile on the distribution shown in Table 32. This table shows the percentile distribution of runs for each run length, taking this on a cumulative percentage, we can identify for the tails of the distribution. I consider anything less than 2.5% of the bottom cumulative percentile as the tail of the distribution at the lower end. I also consider anything more than 2% of the top cumulative percentile as the tail of the distribution at the upper end. Using these definitions, runs of length 1-4 are considered normal and runs of length 5 and above are considered to be in the tail. If we look at the Unbanned Portfolio during the non-ban period, we see that there are 4 runs in the tail of the distribution. Similarly, if we look at the Banned Portfolio during the non-ban period, we see that there are again 4 runs in the tail of the distribution. Both portfolios show similar characteristics in regard to the tails of their distributions.

If we move onto the short sale ban period, we immediately notice that the number of runs in each portfolio is smaller, this is not an effect of the short sale ban but because the relatively short time period of the short sale ban as opposed to the non-ban period. The short sale ban was effective for a period of approximately 3 months, while this is smaller than the approximately 9-month length of the non-ban period. Again, we form a normal distribution in run returns for the short sale ban period for both portfolios.

Looking at the tails of the distributions in the short sale ban period, we see that the Unbanned Portfolio has 1 run in the tail of the distribution, while the Banned Portfolio has 3 runs in the tail of the distribution. We immediately see that the Banned Portfolio is exhibiting fatter tails compared to the Unbanned Portfolio, one of the indications that a short sale ban may be inhibiting the price discovery process.

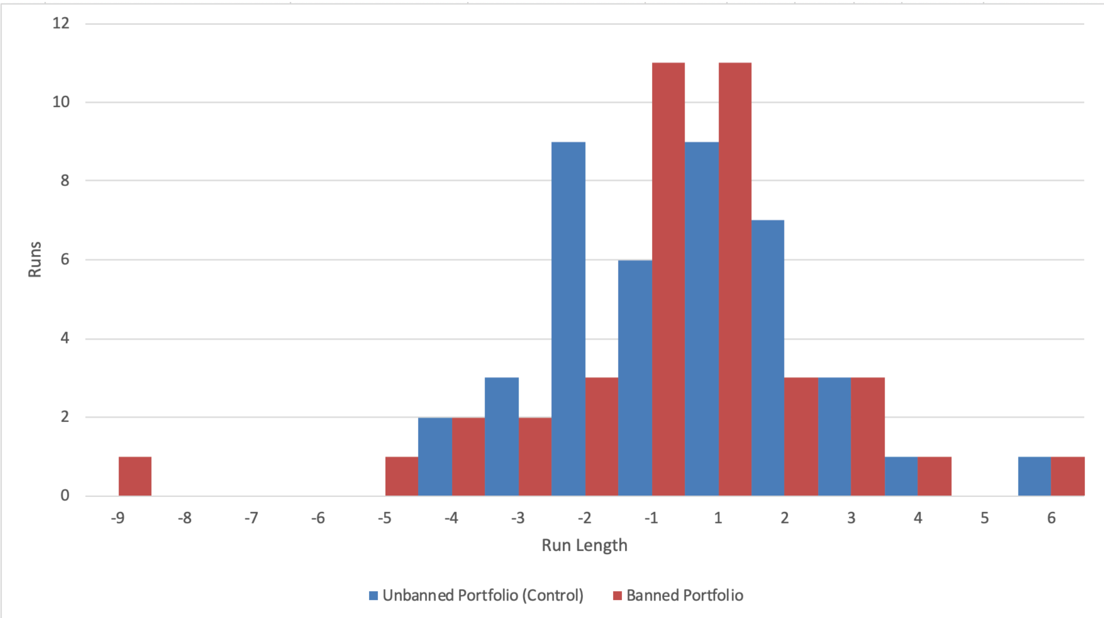
The greater frequency of longer runs in the Banned Portfolio compared to the Unbanned Portfolio in the short sale ban period is indicative of a market needing longer to digest information, whether that is positive or negative news. This in turn is a negative for the market and means new information can't be incorporated into prices as efficiently as we would like.

Figure 28: Banned Portfolio and Unbanned Portfolio Runs Distribution During Non-Ban Period



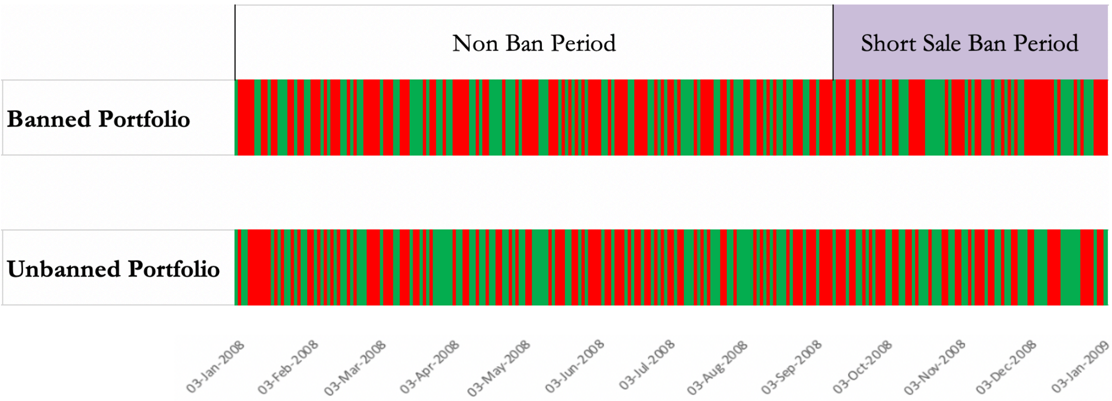
Notes: Figure 28 shows the distribution of runs for the Unbanned Portfolio and Banned Portfolio for the non-ban period for the 2008 Financial Crisis Short Sale Ban in the United Kingdom. The non-ban period includes daily return observations from 2nd January 2008 to 18th September 2008. A run constitutes a consecutive sequence of either positive or negative daily returns for each portfolio.

Figure 29: Banned Portfolio and Unbanned Portfolio Runs Distribution During Short Sale Ban Period



Notes: Figure 29 shows the distribution of runs for the Unbanned Portfolio and Banned Portfolio for the Short Sale Ban Period for the 2008 Financial Crisis Short Sale Ban in the United Kingdom. The Short Sale Ban Period includes daily return observations from 19th September 2008 to 16th January 2009. A Run constitutes a consecutive sequence of either positive or negative daily returns for each portfolio.

Figure 30: Banned Portfolio and Unbanned Portfolio Runs During Non-Ban Period and Short Sale Ban Period.



Notes: Figure 30 shows the Runs in the Unbanned Portfolio and Banned Portfolio for the period where there is not a short sale ban and for the period where there is a short sale ban. A Run constitutes a consecutive sequence of either positive or negative daily returns for each portfolio. A zero-percentage return constitutes the run carrying on until there is a sign change. A red section indicates a negative run period and a green section indicates a positive run period. The thickness of the red and green sections indicates the length of the run periods, where the larger the thickness the longer the run period. 264 daily observations for returns are taken from 3rd January 2008 to 16th January 2009.

Figure 30 shows the distribution of runs in the Banned Portfolio and Unbanned Portfolio across the entire dataset. This graph is unique in the sense that it shows not only the length of the runs with the thickness of the coloured sections (red representing a negative run and green representing a positive run), it also shows the location of the runs during the entire dataset.

We see that the Banned Portfolio generally has larger runs than the Unbanned Portfolio, showing that the Banned Portfolio is much more volatile intraday compared to the Unbanned Portfolio. This however does not account for interday volatility, as a run will only show whether there is zero return, a positive return or a negative return, We take into account that runs are measured over days and volatility is daily, so we can't say that longer run lengths indicate more volatility (as volatility is usually measured on a daily basis) The distribution of runs in the Unbanned Portfolio is very much even, with very little difference in distribution between the non-ban period and short sale ban period. However, for the Banned Portfolio, there is a slight widening in distribution during the short sale ban period over the non-ban period.

As we can see from Figure 30 run patterns are statistically different as well, as well as being observationally different to the eye. We can make a statistical observation from this by comparing percentages of runs between the non-ban and short sale ban period. A percentage of runs would show the percentage of total runs that are of that length in that time period. There are a total of 264 daily observations (3rd January 2008 to 16th January 2009) for our time period, the short sale ban encompasses 83 days (19th September 2008 to 16th January 2009) of those and the normal non ban period encompasses 181 days (3rd January 2008 to 18th September 2018). So, by dividing the number of runs in the non-ban period by 181 and multiplying by 100, we can obtain the percentage run length. The same can be done for the short sale ban period by dividing the number of runs by 83 and multiplying by 100.

Table 32: Unbanned Portfolio and Banned Portfolio Percentage Run Length in Non-Ban Period and Short Sale Ban Period

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Period	Portfolio	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6
Non-Ban Period	Unbanned Portfolio (Control)	0	0	0.552	0	0	1.657	3.315	6.630	17.680	18.785	7.182	2.210	0	1.105	0.552
	Banned Portfolio	0	0	0	0	2.210	2.762	3.315	6.077	11.602	12.707	8.840	2.210	1.657	0	0
Short Sale Ban Period	Unbanned Portfolio (Control)	0	0	0	0	0	2.410	3.614	10.843	7.229	10.843	8.434	3.614	1.205	0	1.205
	Banned Portfolio	1.205	0	0	0	1.205	2.410	2.410	3.614	13.253	13.253	3.614	3.614	1.205	0	1.205

Notes: Table 32 shows the percentage run length for the Unbanned Portfolio and Banned Portfolio for both Non-Ban Period and Ban Period for the 2008 Financial Crisis Short Sale Ban in the United Kingdom. Non-ban period includes daily returns from 2nd January 2008 to 18th September 2008 and the short sale ban period includes daily returns from 19th September 2008 to 16th January 2009. A run constitutes a consecutive sequence of either positive or negative daily returns for each portfolio. A zero-percentage daily return constitutes the run carrying on until there is a sign change. The percentage run length shows the percentage of runs of each length in each time period.

For Table 32, if we focus on the tails of the distribution, runs of length -5 or lower and runs of length 5 or higher, we are able to see a statistical difference in run lengths over the non-ban period and short sale ban period. For the Banned Portfolio, runs of -5 and lower constitute 2.21% of the distribution in the non ban period and 2.41% of the distribution in the short sale ban period. For the Unbanned Portfolio, runs of -5 and lower constitute 0.552% in the non ban period and 0% in the short sale ban period. For the Banned portfolio, runs of 5 and higher constitute 0% in the non ban period and 1.205% in the short sale ban period. For the Unbanned Portfolio, runs of 5 and higher constitute 1.657% in the non ban period and 1.205% in the short sale ban period.

In total for the Banned Portfolio, runs of -5 and lower and 5 and higher make up 2.21% of the distribution in the non ban period and 3.615% of the distribution in the short sale ban period. In total for the Unbanned Portfolio, runs of -5 and lower and 5 and higher make up 2.209% in the non ban period and 1.205% in the short sale ban period. Therefore we see an increase in the runs distribution percentages for the Banned Portfolio over the Unbanned Portfolio.

Table 33 extends the main results further by splitting the Banned Portfolio into 3 distinctive portfolios and replicating the results of Table 31. This allows us further analysis into the behaviour of the Banned Portfolio. The Banned Portfolio consists of financial stocks and I split the stocks in the portfolio into 3 categories of Asset Management, Insurance and Banking depending on the stock's main line of business and revenue stream. **Figure 31** shows a graphical representation of the Asset Management, Insurance and Banking runs distribution for the non-ban period. **Figure 32** shows a graphical representation of the Asset Management, Insurance and Banking runs distribution for the short sale ban period. Both Figure 31 and Figure 32 are graphical representations of the results in Table 32.

The Asset Management Portfolio consists of the firms St James's Place and Standard Life Aberdeen. The Insurance Portfolio consists of the firms RSA Insurance Group, Prudential, Legal and General, Aviva and Admiral Group. The Banking Portfolio consists of the firms Royal Bank of Scotland Group, HSBC Holdings, Standard Chartered, Barclays and Lloyds Banking Group.

Table 33: Asset Management Portfolio, Insurance Portfolio and Banking Portfolio Run Length in Non-Ban Period and Short Sale Ban Period

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Period	Portfolio	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6
Non-Ban Period	Asset Management Portfolio	0	0	0	0	0	0	0	5	2	6	12	23	26	13	6	2	1	0
	Insurance Portfolio	0	0	0	0	0	0	0	4	5	7	10	22	27	14	6	1	0	0
	Banking Portfolio	1	0	0	1	0	0	0	2	5	5	13	15	24	10	3	4	1	0
Short Sale Ban Period	Asset Management Portfolio	0	0	0	0	0	0	0	0	4	3	3	12	12	5	4	0	0	1
	Insurance Portfolio	0	0	0	0	0	0	0	1	1	3	4	14	10	8	3	1	1	0
	Banking Portfolio	0	0	0	0	0	0	1	1	3	1	7	6	7	9	1	1	1	0

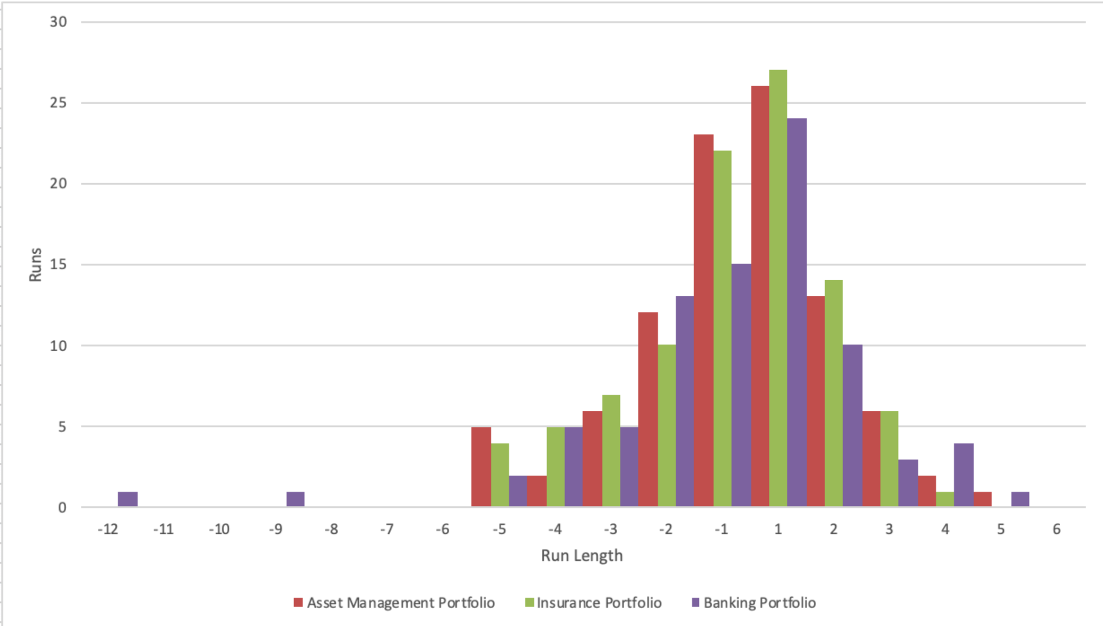
Notes: Table 33 shows the runs for the Asset Management Portfolio, Insurance Portfolio and Banking Portfolio for both Non-Ban Period and Ban Period for the 2008 Financial Crisis Short Sale Ban in the United Kingdom. Non-Ban Period includes daily returns from 2nd January 2008 to 18th September 2008 and the Short Sale Ban Period includes daily returns from 19th September 2008 to 16th January 2009. Portfolios are created from the main revenue stream of either asset management, insurance or banking using stocks in the Banned Portfolio. A run constitutes a consecutive sequence of either positive or negative daily returns for each portfolio. A zero-percentage daily return constitutes the run carrying on until there is a sign change.

In **Table 33** we see for the non-ban period, that the Banking Portfolio exhibits much fatter tails than the Asset Management Portfolio and Insurance Portfolio. All portfolio during the non-ban period exhibit qualities of a normal distribution in their run distribution. If we look at in particular Columns (11), (12) and

(13) both Asset Management Portfolio and Insurance Portfolio have very similar runs distributions. This however is not the case with the Banking Portfolio, which has a much wider distribution of runs compared to the other two.

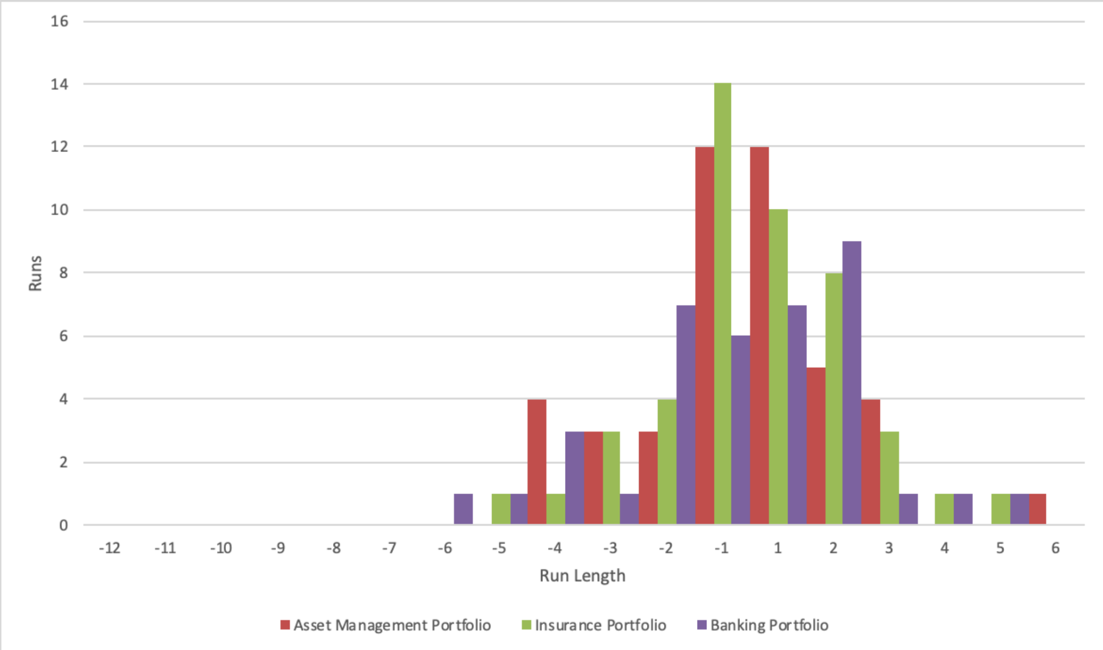
During the short sale ban period, the runs distribution in the Banking Portfolio is much wider than the runs distribution in the Asset Management Portfolio and Insurance Portfolio. It seems a great deal of the fat tails are being generated by the banking stocks, over the asset management and insurance stocks. However, all three categories of stock are contributing to the fat tails.

Figure 31: Asset Management Portfolio, Insurance Portfolio and Banking Portfolio Runs Distribution During Non-Ban Period



Notes: Figure 31 shows the distribution of runs for the Asset Management, Insurance and Banking Portfolio for the Non-Ban Period for the 2008 Financial Crisis Short Sale Ban in the United Kingdom. The Non-Ban Period includes daily return observations from 2nd January 2008 to 18th September 2008. A run constitutes a consecutive sequence of either positive or negative daily returns for each portfolio.

Figure 32: Asset Management Portfolio, Insurance Portfolio and Banking Portfolio Runs Distribution During Short Sale Ban Period



Notes: Figure 32 shows the distribution of runs for the Asset Management, Insurance and Banking Portfolio for the Short Sale Ban Period for the 2008 Financial Crisis Short Sale Ban in the United Kingdom. The Short Sale Ban Period includes daily return observations from 19th September 2008 to 16th January 2009. A run constitutes a consecutive sequence of either positive or negative daily returns for each portfolio.

For the main results, I have initially employed two models, an GARCH (1,1) and Bid-Ask Spread Model and augmented them with dummy variables to see changes between short sale ban and non-ban periods. We see that volatility remains consistent between the two periods, but liquidity deteriorates in banned stocks more than unbanned stocks. Short sales have a significant impact on the market on liquidity from what my study has shown, but little to no effect on volatility. I have also conducted a runs test to see changes in price discovery between the short sale ban period and non-ban periods. We see a deterioration in price discovery in the Banned Portfolio, while the Unbanned Portfolio does not exhibit this deterioration. As an investor the Banned Portfolio has underperformed the Unbanned Portfolio, the Banned Portfolio has experienced greater volatility, deterioration in liquidity and deterioration in price discovery through my study period. These findings are in line with the literature, showing the negative effects of short sale bans overall.

Breaking down the Banned Portfolio further, we have seen that banking stocks are some of the most affected by the short sale ban, while asset management stocks are some of the least affected. Overall, no subsection of the Banned Portfolio is unaffected by the short sale ban. I believe banking stocks are the most effected since these stocks were of the poorest quality in the 2007-2009 financial crisis, the financial crisis stemmed from sub-prime lending where a lot of these banking stocks held their poor-quality loans. The banning of short sales seems to affect the poorer quality stocks the most and the higher quality stocks the least.

6.5 Conclusion

We have seen the results of my research questions. It has been evident from my findings that lightly shorted stocks outperform heavily shorted stocks. Lightly shorted stocks also experience far less volatility and hold stronger balance sheets as a whole. This outperformance of lightly shorted stocks over heavily shorted stocks has been documented from 2001 to 2017 in the US stock market and from 2010 to 2017 in the Canadian stock market. My results are consistent in this regard with Desai et al. (2002) and Boehmer et al. (2010). Lightly shorted stock portfolios are also very much positively correlated with each other in terms of returns, as are heavily shorted stock portfolios with each other in terms of returns. The largest negative correlations are with long/short portfolios and standard portfolios.

What we have also observed is that both on a risk adjusted return and raw return basis, the most lightly shorted stocks perform the best, i.e. the SIR 5% portfolios. Though the distributions of these returns are not strictly uniform[81] across the short interest portfolios. By simply buying, holding and rebalancing the SIR 5% portfolio monthly an investor can achieve an excess return both shown by the Fama and French Three Factor Model with Momentum and the Fama and French Five Factor Model. We saw that the short component of the SIR 95% portfolio was not great enough to hold as part of a strategy and that the best performance could be found by going long the SIR 5% portfolio. An ETF (exchange traded fund) of sort based on short interest data could be a great addition for retail investors if they decided to adopt this strategy.

It is worth noting that by holding any short component, borrowing fees can be occurred in addition. This means portfolio returns for long/short portfolios would be more negative than seen in the results. This is worth considering. Therefore, it is advised with caution in holding a short strategy as costs can quickly rise from borrowing costs and transaction fees as seen in Lu et al. (2018). This is in particular true for high short interest stocks where the cost of borrowing may be higher due to demand to short these stocks.

I further extended these results by conducting a 24-month rolling regression to see changes in alpha (abnormal returns) and beta (volatility with respect to the market) over benchmark models. I find that non-long/short alphas are relatively consistent across time periods, however there are times when this consistency breaks down, this is usually during periods where the market is doing well and there is less risk aversion. Heavily shorted stocks hold lower alphas over lightly shorted stocks with the 24-month rolling regression. I find that betas remain relatively consist across both heavily and lightly shorted portfolios, however after May 2015 the betas for the lightly shorted portfolios drop off, while the betas for the heavily shorted portfolios do not. A possible explanation for this drop off could be other factors such as the SMB or HML of the respective asset pricing models contributing to the pricing of lightly shorted portfolios.

We also saw the effects of short sale bans on volatility, liquidity and price discovery. While volatility was not affected by the short sale ban on either banned or unbanned portfolio, liquidity and price discovery were affected in the banned portfolio. This again highlights the ineffective nature of short sale bans, as shown in

81 As it is not guaranteed that a SIR 5% portfolio will outperform a SIR 10% portfolio over a shorter period of time of a few months, however generally across the broader spectrum, a SIR 5% portfolio is highly likely to outperform a SIR 95% portfolio over a period of several years and decades.

this study. This also brings into doubt market efficiency, as a lack of liquidity can contribute to less efficient pricing of securities. If markets are to be somewhat efficient, an ample supply of liquidity is necessary to counteract overpricing and under-pricing. Short sellers provide a buffer to counteract overpricing, which if withheld can lead to the formation of bubbles and greater more volatile losses further down the line.

I found a good fit for the Bid-Ask Spread Model used to measure the change in liquidity based on R Square and Adjusted R Square results. The bid-ask spread remains stable around 1.8 pence for both portfolios, however once the short sale ban takes place the bid-ask spread increases in both portfolios, but more so in the Banned Portfolio.

I further extended the results by splitting the Banned Portfolio into three portfolios of asset management, insurance and banking to further gauge where in particular the effects of volatility, liquidity and price discovery are seen. In particular I see that the banking stocks behave differently to the asset management stocks and insurance stocks in terms of volatility and price discovery. Banking stocks are seen to be some of the most volatile and the worst affected in price discovery. All portfolios are affected negatively in terms of liquidity, with insurance stocks the worst hit, closely followed by banking stocks. Overall asset management stocks fair the best from the short sale ban compared to banking and insurance stocks, again showing higher quality stocks suffering less from the short-sale ban. However, compared to stocks in the Unbanned Portfolio, the asset management stocks fair far worse. This concludes my findings for my research questions.

CHAPTER 7: CONCLUSION

I started this PhD with the aim of providing insight between short interest and its relationship between market returns, volatility and market manipulation. In particular I set out to answer the following four questions and in turn get a greater understanding of my topics of interest through these four questions. As I have stated at the start of my PhD, my research questions were the following:

1. In adjusting for risk, Boehmer et al. (2010) uses the Fama and French (1993) three-factor model augmented by the momentum factor. However, is this long/short strategy still valid if a different and more recent model, such as Fama and French (2015) five-factor model, is used to adjust for risk premium?
2. After the publication of Boehmer et al. (2010), does the opportunity for excess returns remain, or have investors adopted this strategy and arbitraged away the excess returns?
3. Is the long/short strategy of Boehmer et al. (2010) still valid in an international OECD market that of Canada?
4. Whether and to what extent short sales affect liquidity, price discovery, volatility and cross-section of stock returns?

The first three questions of my project are very much interlinked, as are the three subsections of the last question. The first three questions deal with the relationship between short interest and market returns, whether or whether not a relationship exists and is the relationship consistent across markets. The last question deals with the implications of short selling on market metrics such as liquidity, price discovery, volatility and whether short sellers are trading on good information, noise or bad information.

Research Question 1 looked at whether using a new model for asset pricing such as the Fama and French Five Factor Model meant that the findings of Boehmer et al. (2010) kept robust. Generally, although with some criticism, the Fama and French Five Factor model has seen to be a superior asset pricing model to previous asset pricing models in the literature such as the Capital Asset Pricing Model (CAPM), Fama and French Three Factor Model and the Carhart Four Factor Model. This question looked at whether this new asset pricing model would produce a risk adjusted alpha and show the overperformance of lightly shorted stocks over heavily shorted stocks and thus either be able to produce a long/short strategy with heavily and lightly shorted stocks or a long only strategy with lightly shorted stocks.

I found support for the later, where going long the least shorted portfolio produced both higher returns on a raw basis and held a higher risk adjusted alpha, showing overperformance of lightly shorted stocks when accounting for the Fama and French Five Factor model factors of market risk premium, investment, profitability, the small firm effect and price to book effect.

We saw that there was indeed good news in short interest when the Fama and French Five Factor Model was used to adjust for risk premium. I was not surprised by this result, as we saw that many of the factors overlapped with the Carhart Four Factor Model, I was however surprised by the performance of the heavily shorted stocks as the performance was not as bad as been seen in previous studies in the past. Boehmer et al. (2010) employed a long/short strategy with a relatively weak short component that in my case has turned out to be positive, therefore I would advise only a long strategy of the SIR 5% portfolio.

Research Question 2 built up from Research Question 1 and looked at whether after the publication of Boehmer et al. (2010) the strategy of going long lightly shorted stocks and short heavily shorted stocks remain valid or whether market participants had arbitraged this strategy so that it no longer held true. This question in particular helps contribute to the efficient markets debate, if a market is truly efficient all known public information should be incorporated into a price of a security. I however believe markets are somewhat efficient and are very susceptible to short term overpricing and under-pricing of securities, due to investor psychology and trades based on emotions rather than fundamentals.

I found that again heavily shorted stocks underperformed lightly shorted stocks and the best strategy was to hold the SIR 5% portfolio. It is clear that arbitrage has not taken place and the underperformance of heavily shorted stocks is very much still evident. There could be several reasons for this, most importantly being the publication of Boehmer et al. (2010) in an academic journal over general financial newspapers and magazines such as FT, Forbes or Bloomberg Businessweek. Academic finance does not have the investor capture that a magazine such as Bloomberg Businessweek would have, so many people may still not be aware of such a strategy and its effectiveness. A larger question of interest is why established hedge funds have not adopted this strategy and this very well could be due to other strategies taking priorities or the borrowing costs of shorting securities and rebalancing every month. A fee is usually associated with borrowing a security to short and this price depends on the market, where stocks which have fallen a lot in the short term are usually cheaper to short than stocks which have increased in value a lot in the short term.

Both on a risk adjusted alpha and a raw return basis the best portfolio to hold was the SIR 5% portfolio, I again felt a short component would hinder returns due to the better than expected performance of highly shorted stocks. Highly shorted stocks performed a lot worse than lightly shorted stocks, but not to the extent that they yielded a negative return.

In both Research Question 1 and Research Question 2 the focus had been on post 20th century data in the United States Markets. There is a great deal of historic data on United States securities and its financial system. The United States often sets the standard in both the regulation and infrastructure in financial markets, and thus data is of great depth and availability. However other markets now do exist and are growing at an ever-increasing pace compared to the United States. The academic finance community understand the need to conduct finance research outside the United States and believe this study is one of a limited amount of studies to test short interest data and returns on a country outside the United States.

Short interest data as a dataset can often be very difficult to get hold of and is usually on a paid subscription basis. The large majority of markets in general do not publish short interest data or publish it for a limited number of stocks at infrequent time intervals[82]. I decided to focus on the Canadian markets for Research

82 Finding short interest data on a daily basis is currently very difficult to obtain for all markets, though rebalancing portfolios on a daily basis would be computationally possible by hedge funds to see the true effects of short interest strategies.

Question 3 as returns and short interest data were available from the Toronto Stock Exchange.

I found again heavily shorted stocks underperformed lightly shorted stocks, the performance was more volatile in nature due to the smaller datasets and I was aware of this risk. Again, I found the SIR 5% portfolio to be the best performer and a long only strategy of that portfolio to be the best. It is worth noting that a short component in a long/short strategy can occur borrowing costs, making returns smaller than we have identified. This must also be taken into consideration when applying a short component to any strategy. In particular we must be aware of excessively high borrowing costs (above 50% of stock price in some cases) of stocks with the highest short interest,

What I have noticed over the first three questions is the impact portfolios based on short interest have on returns, there is a clear outperformance of lightly shorted stocks over heavily shorted stocks. Whenever I have tested returns data post 2001 in both the American and Canadian markets, heavily shorted stocks in general have been much more volatile and have underperformed lightly shorted stocks. The reasoning for the outperformance I believe comes down to balance sheet stability and less speculators in lightly shorted stocks over heavily shorted stocks. Speculators contribute a great deal to the volatility.

I further extended Research Question 1,2 and 3 by conducting a 24-month rolling regression to look at alpha and beta performance across my study periods. Alpha is the over or under performance of a portfolio compared to set benchmark such as an asset pricing model return such as the Fama and French Five Factor Model or the Fama and French Three Factor Model with Momentum. Beta is the volatility of a portfolio compared to the market. The benefit of a rolling regression is that it allows you to see changes in a variable across your entire study period, showing how consistent the alpha across the entire study period is. In statistical terms it can be seen as the mean of the alpha on a month by month basis.

I find that heavily shorted portfolios hold consistently lower alphas than lightly shorted portfolios, however there are several periods of time where the divergence in alphas narrows and then rewidens. Usually the narrowing of alphas occurs during periods when the market is seen to be overbought and investors are willing to hold poor quality equities in the anticipation of higher returns over the risk-free rate. It must be noted that the greatest destructions of capital are noted when investors hold poor quality equities in times of good economic conditions. This often leads to overvaluation in these equities which is corrected when the next bear market occurs and credit tightening leads to financial hardships for these companies. Investors can risk overpaying for a high-quality asset, but overpaying for a poor-quality asset is dangerous and often leads to destruction of capital. Regarding betas, all portfolios were highly correlated in their betas, however after May 2015 this correlation broke down. This was observed for both United States and Canadian markets. A possible reason for this breakdown could be another factor of the asset pricing model having more influence on the returns of the lightly shorted portfolios.

Going forward with the research in these first three questions there is a lot of scope for development, many stock markets such as the UK are now developing better databases of short interest data and thus the Boehmer et al. (2010) findings can be tested across these markets as well. Returns data in general is not very hard to find and can often be constructed from closing prices, however short interest data is often subscription specific, so can be very expensive to get hold of in particular on an individual basis. With the access of financial markets to more and more retail traders as opposed to institution traders, short interest data is becoming more and more valuable to both short sellers and those taking a short-term long position in a security. I hope because of this; more and more short interest databases are made accessible to the public and universities in general.

There is scope for my techniques to be automated via computer coding designed to construct portfolios based off short interest data and back test results. This sort of automated technique can also be applied to other characteristics of securities as well and is not just limited to short interest. Portfolios based on P/E ratios or P/B ratios can be constructed in a similar way, so can portfolios based on dividend yield or dividend cover[83]. If we are able to establish a clear-cut link with outperformance of a certain criteria of stock over others, we may be able to create better models for judging the return of a security compared to the ones we have today. The P/E portfolio back testing would be a classic matchup between value and growth stocks. The dividend yield back testing would be a good measure of a company's ability to effectively employ retained earnings compared to giving it to investors. A company paying little to no dividend has the ability to take earnings and either employ the capital in other businesses or buy back their own stock at sensible valuations. An underperformance of non-dividend paying stocks may possibly show a destruction of capital by past management teams.

Another point to mention is that over the first three questions, I conducted the back testing using short interest data published on a monthly basis, I am now aware databases are being constructed on a weekly and even daily basis. This again, though very time consuming if not automated, would give an even clearer link between short interest and returns. Stocks can change short interest on a month by month basis as we have seen and in particular this is exaggerated in small cap illiquid stocks, taking short interest data on a daily basis would definitely aid in research in those small cap illiquid stocks. I believe that large cap stocks with sustained high volumes are less volatile to changes in short interest especially over the period of a few days, so it may not be as beneficial for them.

Regarding research question 4, I looked to investigate the relationship between short selling and the affect it has on the liquidity, volatility and price discovery of stocks. Short selling is naturally implemented in most markets in order to allow efficient markets and better price discovery, it means overvalued stocks are brought down and undervalued stocks are brought up from shorts closing their position. The extra volume adds to efficiency in the market and efficient markets are ideal for both investors, financial regulatory authorities such as the FCA or the SEC and central banks such as the US Federal Reserve or the Bank of England.

To investigate how volatility affects stocks, I used the 2008 short sale ban as a proxy. I used a GARCH (1,1) model with a dummy variable to see how volatility changes between no short sale ban and short sale ban periods. An GARCH model is ideal as it exhibits the quality of volatility clustering, where high and low periods of volatility remain distinct. An EGARCH model would have been applied for the main methodology if all the data had shown elements of asymmetric returns behaviour to stock news, however

[83] Dividend cover looks at the number of times the dividend can be covered from earnings, this shows us the sustainability of the dividend. A dividend cover below 1 usually indicates the current dividend is at risk and will most likely be cut or suspended unless earnings rise. The natural way of increasing the dividend cover is to increase earnings per share (EPS) either via higher earnings or with stock buybacks.

this was not the case. The dummy variable meant we could see whether volatility changed between the two periods and we could look at significance levels to see if the change was meaningful. I employed two portfolios, one investigating stocks that were banned and another investigating stocks that were not banned. The portfolio with unbanned stocks acted as a control variable.

I found no change in volatility between the two periods being investigated, further showing whether a short sale ban is effective at all. Both dummy variable coefficients were not significant, indicating that the control portfolio also experienced no change in volatility. A short sale ban is implemented on the reason that downward pressure on stocks can be reduced and volatility lessened, we saw a resumption of downward pressure on financial stocks and lack of change in volatility. My study like many others questions whether short sale bans have any meaningful impact. A lot of firms often blame depressed stock prices because of short sellers, but fairer prices can be produced with the upward and downward pressure short sellers are able to impact on the price of a stock.

Regarding model fit it was seen that the GARCH (1,1) model fit the Unbanned portfolio and Banned Portfolio very well, the banned portfolio of stocks showed much less asymmetry than the unbanned portfolio of stocks. This led me to the conclusion that an EGARCH model could have been a good fit for the Unbanned Portfolio. However, the fit for the Banned Portfolio would be poor and a GARCH specification would suffice. It is difficult to find a model that is unique and that fits two sets of data as well. It would not have been good to use two different models to answer one question, as it makes comparison between coefficients much more difficult, however running extra models in the appendix is a solution. Model fitting is never perfect, though AIC and BIC criterion mean that we can be certain that we are using the best model that we have in that family of models.

To investigate liquidity, I again use the 2008 short sale ban as a proxy. Bid-ask spreads are often used as a gauge of the liquidity in a security. The wider the bid-ask spread, the more illiquid a stock is said to be. It can often be seen that a stock with a wider bid-ask spread is riskier to trade, as the security needs to move more than the spread and fx and/or trading fees in order to make a profit for the trader. I find that liquidity deteriorates in both Banned and Unbanned Portfolios when the short sale ban takes place, however the deterioration in liquidity is much more evident in the Banned Portfolio, this can be seen with the wider bid-ask spread in the Banned Portfolio. Overall this shows us that the short sale ban causes a loss of liquidity in securities and thus impacts market efficiency. The Bid-Ask Spread Model also showed a good fit in regards to R Square and Adjusted R Square, again showing coefficients are contributing to the fit of the model. The similarity between R Square and Adjusted R Square shows most, if not all of the coefficients are contributing to the fit of the model.

To investigate price discovery, I use the Wald-Wolfowitz Runs Test to gauge for fat tails in the run distributions of both banned and unbanned stock samples. Fat tails are indicative of short selling bans affecting price discovery. A runs test is an effective means of investigating price discovery when a full list of trade data for individual short sales is not available for the stocks being investigated. I find longer runs in the Banned Portfolio compared to the Unbanned Portfolio, showing fatter tails. This shows us that a deterioration in price discovery has most likely taken place, due to the short sale ban. This is consistent with the existing literature, where short sale bans are seen to affect price discovery in a negative fashion.

I also split the stocks in the Banned Portfolio into three categories of asset management, insurance and banking based on the main revenue stream of these companies. In particular I see that the banking stocks behave differently to the asset management stocks and insurance stocks with respect to volatility and price discovery. Banking stocks are seen to be some of the most volatile and the worst affected in price discovery. All portfolios of asset management, insurance and banking are affected negatively in terms of liquidity, with insurance stocks the worst hit, closely followed by banking stocks. Asset management stocks are affected the least from the short sale ban compared to banking and insurance stocks. However, no stock in the Banned Portfolio behaves as well as the Unbanned Portfolio in terms of market efficiency, volatility and liquidity.

The findings of research question 4 are consistent with many other studies and shows that short sale bans damage the market more than they help it. If the purpose of a short sale ban is to stabilise a market, this is not holding true as key market stabilising metrics such as liquidity and price discovery are being hampered. My strategy going forward would be to not impose any form of short sale ban in a falling market, but to be more accommodative in policy to help restore confidence in a market. A market often trades on emotions and more accommodative language[84] by the Federal Reserve, Bank of England or other central bank can help restore confidence in a market. A short sale ban may decrease extra selling pressure but it will have no impact on investors or speculators who are a currently in a long position of a security from selling.

A trading halt may be more worthwhile over an extended period over a short sale ban. It may be worthwhile in the future looking at the effects of trading halts on stabilising markets compared to outright short sale bans. The deterioration in liquidity is always a concern for a market, as it leads to inconsistencies in pricing, which has been further enhanced by my price discovery findings. Long term investors should not be opposed to a period of closure in the market, which hampers in particular short-term speculators.

In a broader sense it may be of interest of regulators to promote investment over speculation[85], where the vast majority of shorting activity takes place. This could be done by increasing taxes on financial transactions in such a way that short term speculation is not profitable or in turn, giving tax breaks for capital gains in investments held over a longer period of time. Another strategy would be to impose a higher minimum stock price value before delisting, a speculator prefers low priced stocks with ample liquidity as they are able to transact with low levels of capital quickly. A final consideration is to ban the selling and buying of partial stocks, where an investor for example can buy a 1/10th of a share of Amazon or 1/5th a share of Apple. Again, these lower stock prices promote speculation and allow people to bet on the direction of stocks in

---

84 Accommodative language in the sense of language that benefits equity investment over other securities such as gold or cash. This can often mean more dovish language in terms of interest rate increases or means of the central bank hinting at buying bonds or stocks in the open market to support higher market prices. It can also mean supporting financial institution with capital in times of panic.

85 The clear distinction between the investor and the speculator is their behaviour towards stock price movements. A speculator is only concerned with the price of a stock, with little regard to the nature or the fundamentals of a business. The investor is concerned about the nature and fundamentals of a business and is not concerned with the daily fluctuations in the price of a stock. The investor will not be worried if the market is shut for a period of years, as they look towards the workings of a business in the long run. The speculator on the other hand, would be worried if the market is closed for a period of years as this will remove market fluctuations from his or her advantage. Daily stock quotations by the market promote speculation over investment.

the short term, over long term holding.

These are considerations worth taking note of in regards to short sale bans in the market. Of course, these considerations can be implemented during volatile times in a market or for a longer duration. It is a fine balance between maintaining the liquidity from speculators and also driving a market towards long term investment over short term speculation.

REFERENCES

Please find below a full list of references used throughout the thesis in Harvard style. References quoted in text will match these references by author surname and year, however these references are in full and will include author surname and initials, year, title of publication and journal or place of publication. References are listed in alphabetical order by author surname.

Ackert, L.F., Athanassakos, G., 2005. The Relationship Between Short Interest and Stock Returns in the Canadian Market. Journal of Banking and Finance 29, 1729-1749.

Ahn, H., Bae, K., Chan, K., 2001. Limit Orders, Depth, and Volatility: Evidence from the Stock Exchange of Hong Kong. The Journal of Finance 56, 767-788.

Akbas, F., Boehmer, E., Ertruk, B., Sorescu, S., 2008. Why Do Short Interest Levels Predict Stock Returns? Unpublished Working Paper, Texas A&M University.

Alves, C., Mendes, V., Pereira da Silva, P., 2016. Analysis of Market Quality Before and During Short Selling Bans. Research in International Business and Finance 37, 252-268.

Anderson, K., Stafylas, D., Uddin, M., 2018. Hedge Fund Performance Attribution Under Various Market Conditions. International Review of Financial Analysis 56, 221-237.

Asquith, P., Meulbroek, L., 1995. An Empirical Investigation of Short Interest. Unpublished Working Paper, MIT.

Asquith, P., Pathak, P.A., Ritter, J.R., 2005. Short Interest, Institutional Ownership and Stock Returns. Journal of Financial Economics 78, 243-276.

Au, A.S., Doukas, J.A., Onayev, Z., 2009. Daily Short Interest, Idiosyncratic Risk and Stock Returns. Journal of Financial Markets 12, 290-316.

Banz, R.W., 1981. The Relationship Between Return and Market Value of Common Stocks. Journal of Financial Economics 9, 3-18.

Barinov, A., Wu, J., 2014. High Short Interest Effect and Aggregate Volatility Risk. Journal of Financial Markets 21, 98-122.

Bary, E., 2019. These are the Most Expensive Stocks for Short Sellers to Borrow. URL: www.marketwatch.com/story/these-are-the-most-expensive-stocks-for-short-sellers-to-borrow-2019-05-30. Retrieved 1st April 2020.

Basu, S., 1977. Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis. The Journal of Finance 32, 663-682.

Battalio, R., Schultz, P., 2011. Regulatory Uncertainty and Market Liquidity: The 2008 Short Sale Ban's Impact on Equity Option Markets. The Journal of Finance 66, 2013– 2053.

Beber, A., Pagano, M., 2013. Short-selling Bans Around the World: Evidence from the 2007–09 Crisis. The Journal of Finance 68, 343–381.

Black, F., 1972. Capital Market Equilibrium with Restricted Borrowing. The Journal of Business 45, 444-455.

Blitz, D., Hanauer, M.X., Vidojevic, M., Van Vliet, P., 2018. Five Concerns with the Five-Factor Model. The Journal of Portfolio Management Quantitative Special Issue 2018 44, 71-78.

Boehmer, E., Huszar, Z.R., Jordan, B.D., 2010. The Good News in Short Interest. Journal of Financial Economics 96, 80-97.

Boehmer, E., Jones, C.M, Zhang, X., 2008. Which Shorts are Informed? The Journal of Finance 63, 491-527.

Boehmer, E., Wu, J., 2013. Short Selling and the Pricing Discovery Process. Review of Financial Studies 26, 287-322.

Bodie, Z., Kane, A., Marcus, A.J., 2015. Investments Global Edition. McGraw-Hill Education.

Bohl, M.T., Reher, G., Wilfling, B., 2016. Short Selling Constraints and Stock Returns Volatility: Empirical Evidence from the German Stock Market. Economic Modelling 58, 159-166.

Bordo, M.D., Redish, A., Rockoff, H., 2015. Why didn't Canada have a Banking Crisis in 2008 (or in 1930, or 1907, or…)? The Economic History Review 68, 218-243.

Bris, A., Goetzmann, W.N., Zhu, N., 2007. Efficiency and the Bear: Short Sales and Markets Around the World. The Journal of Finance 62, 1029-1079.

Brockman, P., Hao, Q., 2011. Short Selling and Price Discovery: Evidence from American Depositary Receipts. The Journal of Financial Research 34, 569- 588.

Buffett, W.E., 2004. Chairman's 2004 Letter to Berkshire Shareholders.

Callen, J.L., Fang, X., 2015. Short Interest and Stock Price Crash Risk. Journal of Banking and Finance 60, 181-194.

Carhart, M., 1997. On Persistence in Mutual Fund Performance. The Journal of Finance 52, 57-82.

Chan, K.C., Feng, X., Yang, D., 2017. Short Sale Constraints, Dispersion of Opinion and Stock Overvaluation: Evidence from Earnings Announcements in China. The North American Journal of Economics and Finance 41, 217-230.

Chang, E.C., Cheng, J.W., Yu, Y., 2007. Short Sales Constraints and Price Discovery: Evidence from the Hong-Kong Market. The Journal of Finance 62, 2097-2121.

Chen, N., Roll, R., Ross, S.A., 1986. Economic Forces and the Stock Market. The Journal of Business 59, 383-403.

Chiah, M., Chai, D., Zhong, A., Li, S., 2016. A Better Model? An Empirical Investigation of the Fama French Five Factor Model in Australia. International Review of Finance 16, 595-638.

Clark, P., 1973. A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. Econometrica 41, 135-155.

Clifton, M., Snape, M., 2008. The Effect of Short Selling Restrictions on Liquidity: Evidence from the London Stock Exchange. London Stock Exchange Commissioned Paper.

Comerton-Forde, C., Putnins, T.J., 2009. Are Short Sellers Manipulating the Market? Unpublished Paper.

Conaldi, G., 2016. RESE1044 - Research Methods Course Handbook. University of Greenwich Research Methods Resources.

Copeland, T.E., 1976. A Model of Asset Trading Under the Assumption of Sequential Information Arrival. The Journal of Finance 31, 1149-1168.

Corwin, S.A., Schultz, P., 2012. A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices. The Journal of Finance 67, 719-760.

Del Guercio, D., Reuter, J., 2014. Mutual Fund Performance and the Incentive to Generate Alpha. The Journal of Finance 69, 1673-1704.

Desai, H., Ramesh, K., Thiagarajan, S.R., Balachandran, B.V., 2002. An Investigation of the Informational Role of Short Interest in the Nasdaq Market. The Journal of Finance 57, 2263-2287.

De Graaf, H., Kalse, E., 2008. Naakt short gaan, een oud-Hollands kunstje. URL: https://www.nrc.nl/nieuws/2008/07/25/naakt-short-gaan-een-oud-hollands-kunstje-11578528-a1091676. Retrieved 25 July 2019.

Diamond, D., Verrecchia, R., 1987. Constraints on Short Selling and Asset Price Adjustment to Private Information. Journal of Financial Economics 18, 277-311.

Diether, K.B., Lee, K., Werner, I.M., 2009. Short Sale Strategies and Return Predictability. The Review of Financial Studies 22, 575-607.

Drechsler, I., Drechsler, Q.S., 2016. The Shorting Premium and Asset Pricing Anomalies.

Engelberg, J.E., Reed, A.V., Ringgenberg, M.C., 2012. How are Shorts Informed? Short Sellers, News and Information Processing. Journal of Financial Economics 105, 260-278.

Engle, R.F., Ng, V.K., 1993. Measuring and testing the impact of news on volatility. Journal of Finance 48, 1749–1778.

Fama, E.F., French, K.R., 1993. Common Risk Factors in the Return of Stocks and Bonds. Journal of Financial Economics 33, 3-56.

Fama, E.F., French, K.R., 2004. The Capital Asset Pricing Model: Theory and Evidence. Journal of Economic Perspectives 18, 25-46.

Fama, E.F., French, K.R., 2015. A Five Factor Asset Pricing Model. Journal of Financial Economics 116, 1-22.

Figlewski, S., Webb, G.P., 1993. Options, Short Sales, and Market Completeness. The Journal of Finance 48, 761-777.

Floyd, E., Li, N., Skinner, D.J., 2015. Payout Policy through the Financial Crisis: The Growth of Repurchases and the Resilience of Dividends. Journal of Financial Economics 118, 299-316.

Fotak, V., Raman, V., Yadav, P.K., 2009. Naked Short Selling: The Emperor's New Clothes?, Unpublished Paper.

Foye, J., Mramor, D., Pahor, M., 2013. A Respecified Fama French Three Factor Model for the New European Union Member States. Journal of International Financial Management and Accounting 24, 3-25.

Goldstein, I., Guembel, A., 2008. Manipulation and the Allocational Role of Prices. The Review of Economic Studies 75, 133-164.

Graham, B., 1949. The Intelligent Investor. Harper & Brothers.

Griffin, J.M., 2002. Are the Fama French Factors Global or Country Specific? The Review of Financial Studies 15, 783-803.

Griffin, J.M., Harris, J., Shu, T., Topaloglu, S., 2011. Who Drove and Burst the Tech Bubble? The Journal of Finance 66, 1251-1290.

Haruvy, E., Noussair, C.N., 2006. The Effect of Short Selling on Bubbles and Crashes in Experimental Spot Asset Markets. The Journal of Finance 61, 1119-1157.

Harvey, C.R, Graham, J.R, 2001. The Theory and Practice of Corporate Finance: Evidence from the Field. Journal of Financial Economics 60, 187-243.

Hauser, R., 2013. Did Dividend Policy Change During the Financial Crisis? Managerial Finance 39, 584-606.

Hendershott, T., Kozhan, R., Raman, V., 2018. Short Selling and pricing Discovery in Corporate Bonds. Journal of Financial and Quantitative Analysis (In Press).

Ho, K.W., 1996. Short Sale Restrictions and Volatility: The Case of the Stock Exchange of Singapore. Pacific-Basin Finance Journal 4, 377-391.

Hodgkinson, L., Jaafar, A., Mohamad, A., Wells, J., 2013. Short Selling and Stock Returns: Evidence from the UK. The British Accounting Review 45, 125-137.

Howell, E., 2016. Short Selling Restrictions in the EU and the US: A Comparative Analysis. Journal of Corporate Law Studies 333.

Islam, A., Khaled, M., 2005. Tests of Weak-Form Efficiency of the Dhaka Stock Exchange. Journal of Business Finance and Accounting 32, 1613–24.

Jagannathan, R., Wang, Z., 1996. The Conditional CAPM and the Cross Section of Expected Returns. The Journal of Finance 51, 3-53.

Lee, K., Wang, S., 2015. Do Foreign Short Sellers Predict Stock Returns? Evidence from Daily Short Selling in Korean Stock Market. Pacific-Basin Finance Journal 32, 56-75.

Lehmann, B.N., Modest, D.M., 1985. The Empirical Foundations of the Arbitrage Pricing Theory II: The Optimal Construction of Basis Portfolios. The National Bureau of Economic Research.

Lippi, A., Barbieri, L., Piva, M., De Bondt, W., 2018. Time-varying Risk Behaviour and Prior Investment Outcomes: Evidence from Italy. Judgment and Decision Making 13, 471-483.

Lin, Q., 2017. Noisy Prices and the Fama French Five Factor Asset Pricing Model in China. Emerging Markets Review 31, 141-163.

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, 13-37.

Lobanova, O., Hamid, S.S., Prakash, A.J., 2010. The Impact of Short-Sale Restrictions on Volatility, Liquidity and Market Efficiency: The Evidence from the Short-Sale Ban in the U.S. Unpublished.

Liu, Q., Rhee, S.G., Zhang, L., 2015. Limits to Arbitrage in US Penny Stocks: An Overlooked Segment of US Equity Markets. Asian Finance Association 2015 Conference Paper.

Lu, T., Ren, J., Zhao, Y., 2018. Costly Long-Short Strategies Under Short-Sale Constraints: Chinese Evidence. International Review of Finance 18, 743-751.

Mandelbrot, B.B., 1963. The Variation of Certain Speculative Prices. The Journal of Business 36, No. 4, 394-419

Markowitz, H., 1952. Portfolio Selection. The Journal of Finance 7, 77-91.

Marsh, I.W., Payne, R., 2012. Banning Short Sales and Market Quality: The UK's Experience. Journal of Banking and Finance 36, 1975–1986.

Miller, E.M., 1977. Risk, Uncertainty and Divergence of Opinion. The Journal of Finance 32, 1151-1168.

Modigliani, F., Miller, M.H., 1958. The Cost of Capital, Corporation Finance and the Theory of Investment. American Economic Review 48, 261-297.

Mollik, A., Bepari M.D., 2009. Weak-Form Market Efficiency of Dhaka Stock Exchange (DSE), Bangladesh. Unpublished Paper presented at 22nd Australasian Finance and Banking Conference.

Mossin, J., 1966. Equilibrium in a Capital Asset Market. Econometrica 34, 768-783.

Nelson, D.B., 1991. Conditional Heteroskedasticity in Asset Returns: A New Approach. Econometrica 59, 347–370.

Nezafat, M., Shen, T., Wang, Q., Wu, J., 2015. Longs, Shorts and the Cross-Section of Stock Returns. Unpublished Paper.

Novy-Marx, R., 2013. The Other Side of Value: The Gross Profitability Premium. Journal of Financial Economics 108, 1-28.

Rapach, D.E., Ringgenberg, M.C., Zhou, G., 2016. Short Interest and Aggregate Stock Returns. Journal of Financial Economics 121, 46-65.

Reuters, 2008. Short sellers have been the villain for 400 years. URL: https://www.reuters.com/article/us-financial-shortselling-villainspics/short-sellers-have-been-the-villain-for-400-years idUSTRE48P7CS20080926?PageNumber. Retrieved 25 July 2019.

Ripamonti, A., 2016. Corwin-Schultz Bid-Ask Spread Estimator in the Brazilian Stock Market. Brazilian Administration Review 13, 76-97.

Ritter, J.R., Welch, I., 2002. A Review of IPO Activity, Pricing and Allocations. The Journal of Finance 57, 1795-1828.

Roll, R., 1977. A Critique of the Asset Pricing Theory's Tests Part I: On Past and Potential Testability of the Theory. Journal of Financial Economics 4, 129-176.

Roll, R., Ross, S.A., 1980. An Empirical Investigation of the Arbitrage Pricing Theory. The Journal of Finance 35, 1073-1103.

Roll, R., 1984. A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market. The Journal of Finance 39, 1127-1139.

Ross, S.A., 1976. The Arbitrage Theory of Capital Asset Pricing. Journal of Economic Theory 13, 341-360.

Schultz, P. 2003. Pseudo Market Timing and the Long-Run Underperformance of IPOs. The Journal of Finance 58, 483-517.

Sentana, E., 1995. Quadratic ARCH Models. Review of Economic Studies 62, 639-661.

Sharpe, W.F., 1963. A Simplified Model for Portfolio Analysis. Management Science 9, 277-293.

Sharpe, W.F., 1964. Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. The Journal of Finance 19, 425-442.

Shkilko, A., Van Ness, B., Van Ness, R., 2009. Aggressive Short Selling and Price Reversals, Unpublished Paper.

Sirri, E.R., 2009. Regulatory Politics and Short Selling. University of Pittsburgh Law Review 71.

Sochi, M., Swidler, S., 2018. A Test of Market Efficiency when Short Selling is Prohibited: A Case of the Dhaka Stock Exchange. Journal of Risk and Financial Management 11, 59.

Tian, H., 2014. The Effects of a Temporary Short Selling Ban. Unpublished Paper

Treynor, J.L., 1961. Market Value, Time and Risk. Unpublished Paper.

Treynor, J.L., 1962. Towards a Theory of Market Value of Risky Assets. Unpublished Paper.

Ulku, N., 2017. Monday Effect in Fama French's RMW Factor. Economics Letters 150, 44-47.

Zhisheng L., Bingxuan L., Ting Z., Chen, C., 2018. Does short selling improve stock price efficiency and liquidity? Evidence from a natural experiment in China. The European Journal of Finance 24, 1350-1368.

APPENDICES

Appendix 1:

Table A1: Unbanned Portfolio and Banned Portfolio EGARCH (2,3) Model with Dummy Variable

		Unbanned Portfolio (Control)		Banned Portfolio	
		(1)	(2)	(3)	(4)
Equation	Variable	Coefficient	P-value	Coefficient	P-value
Mean Equation	Constant	0.000	(0.505)	0.000	(0.942)
	FTSE 100 daily return	0.733***	(0.000)	1.168***	(0.000)
Conditional Equation	Constant	-0.141	(0.239)	-0.891	(0.227)
	EARCH 1	0.107	(0.150)	0.058	(0.498)
	EARCH 2	0.007	(0.939)	-0.009	(0.917)
	EARCH_A 1	0.501***	(0.000)	0.201	(0.168)
	EARCH_A 2	-0.342**	(0.033)	0.367***	(0.007)
	EGARCH 1	0.888***	(0.000)	-0.294	(0.503)
	EGARCH 2	0.675***	(0.000)	0.935***	(0.000)
	EGARCH 3	-0.578***	(0.001)	0.240	(0.583)
	Dummy	0.001	(0.166)	-0.002	(0.518)

Notes: Table A1 shows the EGARCH (2,3) model with dummy variable results for both Unbanned and Banned Portfolios. I report for both Mean and Conditional Equations. With coefficients I report for the Constant coefficient, FTSE 100 daily return coefficient, EARCH 1 coefficient, EARCH 2 coefficient, EARCH_A 1 coefficient, EARCH_A 2 coefficient, EGARCH 1 coefficient, EGARCH 2 coefficient, EGARCH 3 coefficient and Dummy coefficient. P-values are given in brackets and indicate significance levels. I use a star system like previous to denote significance, *** is significance at the 1% level, ** is significance at the 5% level and * is significance at the 10% level. I investigate banned stocks but use unbanned stocks as a control variable. Data author's own calculation, calculated using STATA 15. The following model is used for the production of these results:

$$R_t = \alpha_t + \beta_i R_{mt} + \varepsilon_t$$

$$\log\varepsilon_t^2 = \omega + \beta_1 g(Z_{t-1}) + \beta_2 g(Z_{t-2}) + \alpha_1 log\sigma_{t-1}^2 + \alpha_2 log\sigma_{t-2}^2 + \alpha_3 log\sigma_{t-3}^2 + D_{SSB}$$

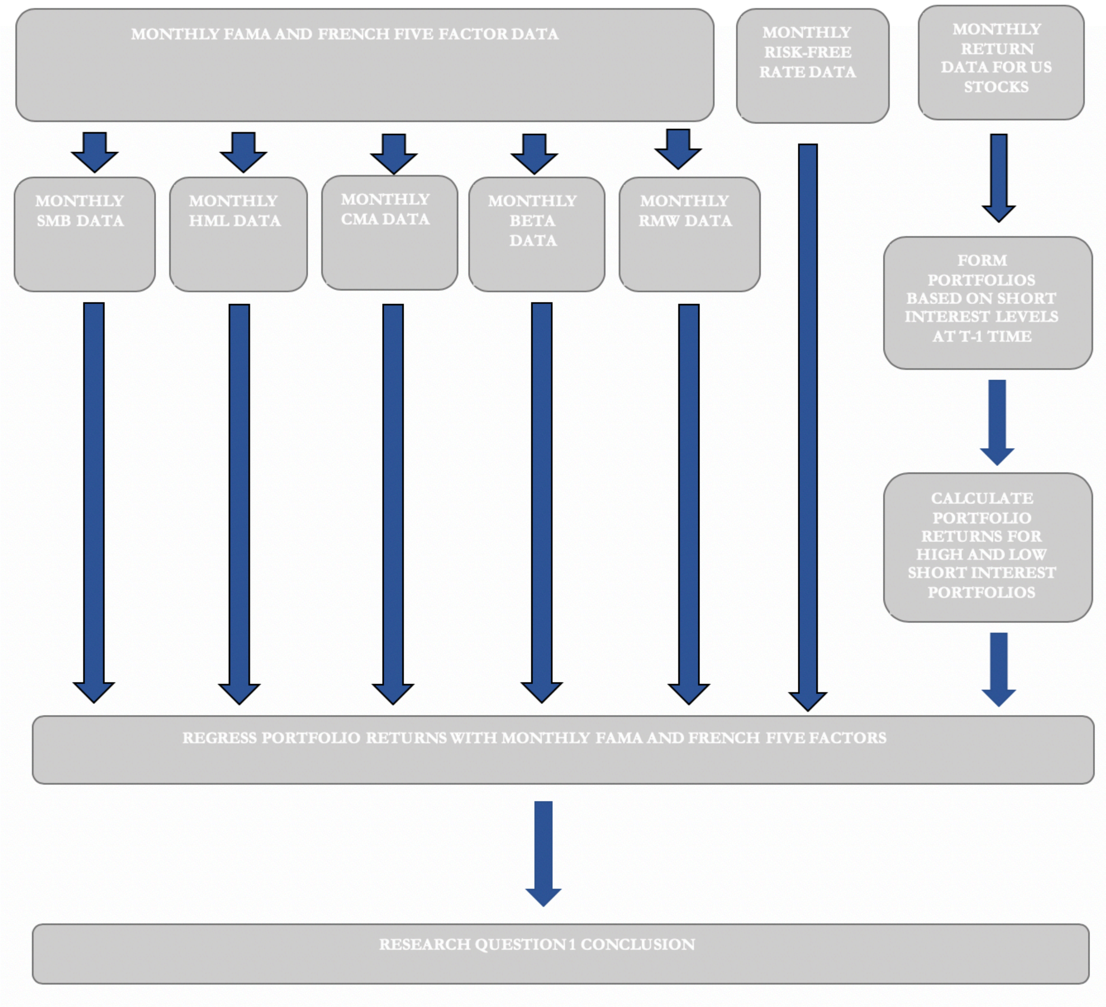
Where:

$$\boldsymbol{g(Z_t) = \theta Z_t + \lambda(|Z_t| - E(|Z_t|))}$$

The model above represents an EGARCH (2,3) model with a dummy variable to take note of changes in volatility between the no short sale ban and short sale ban time periods. R_t is the portfolio return at time t and R_{mt} is the FTSE 100 daily return at time t. Z_{t-2} is the ARCH effect at time t-2, Z_{t-1} is the ARCH effect at time t-1 and Z_t is the ARCH effect at time t. σ_{t-1}^2 is the GARCH effect at time t-1, σ_{t-2}^2 is the GARCH effect at time t-2 and σ_{t-3}^2 is the GARCH effect at time t-3. D_{SSB} is the dummy variable. ε_t is the error term. I take daily observations between January 3rd 2008 and January 16th 2009, total of 264 daily observations.

Appendix 2:

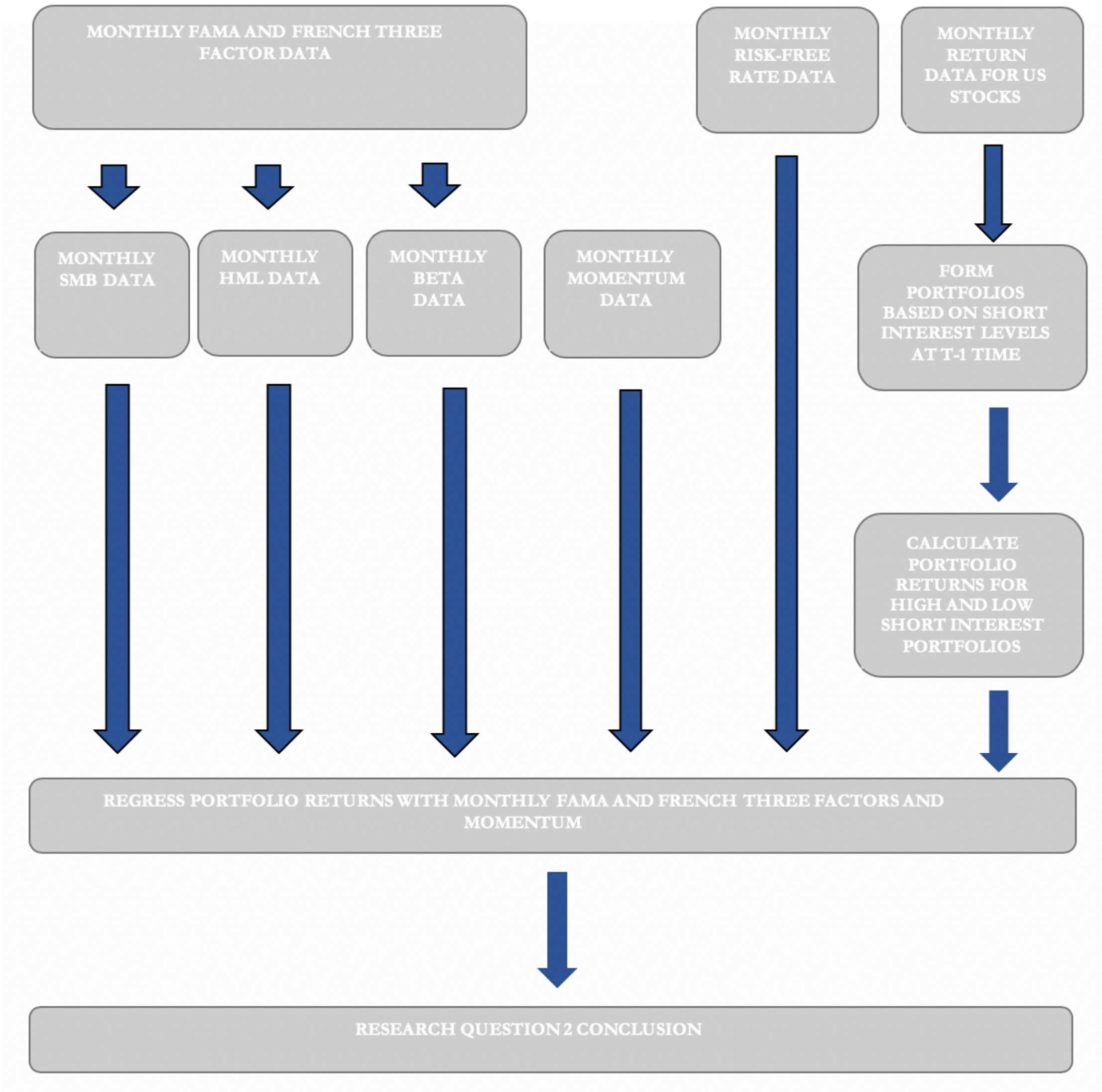
Figure A1: Research Question 1 Methodology Diagram



Notes: Figure A1 shows a methodology diagram outlining the steps that need to be taken in order to produce results for research question 1. The methodology diagram allows you to see clearly what data needs to be collected, what the data will help to produce and how from that I will achieve my conclusion for research question 1. Monthly Fama and French Five Factor data is collected, consisting of the factors of SMB (Small Minus Big), HML (High Minus Low), CMA (Conservative Minus Aggressive), Beta and RMW (Robust Minus Weak). Monthly returns data is collected for US stocks and portfolios are formed on short interest levels at time t-1. Portfolio returns are calculated from this. This is combined with the monthly risk-free rate and excess portfolio returns are regressed against the five factors.

Appendix 3:

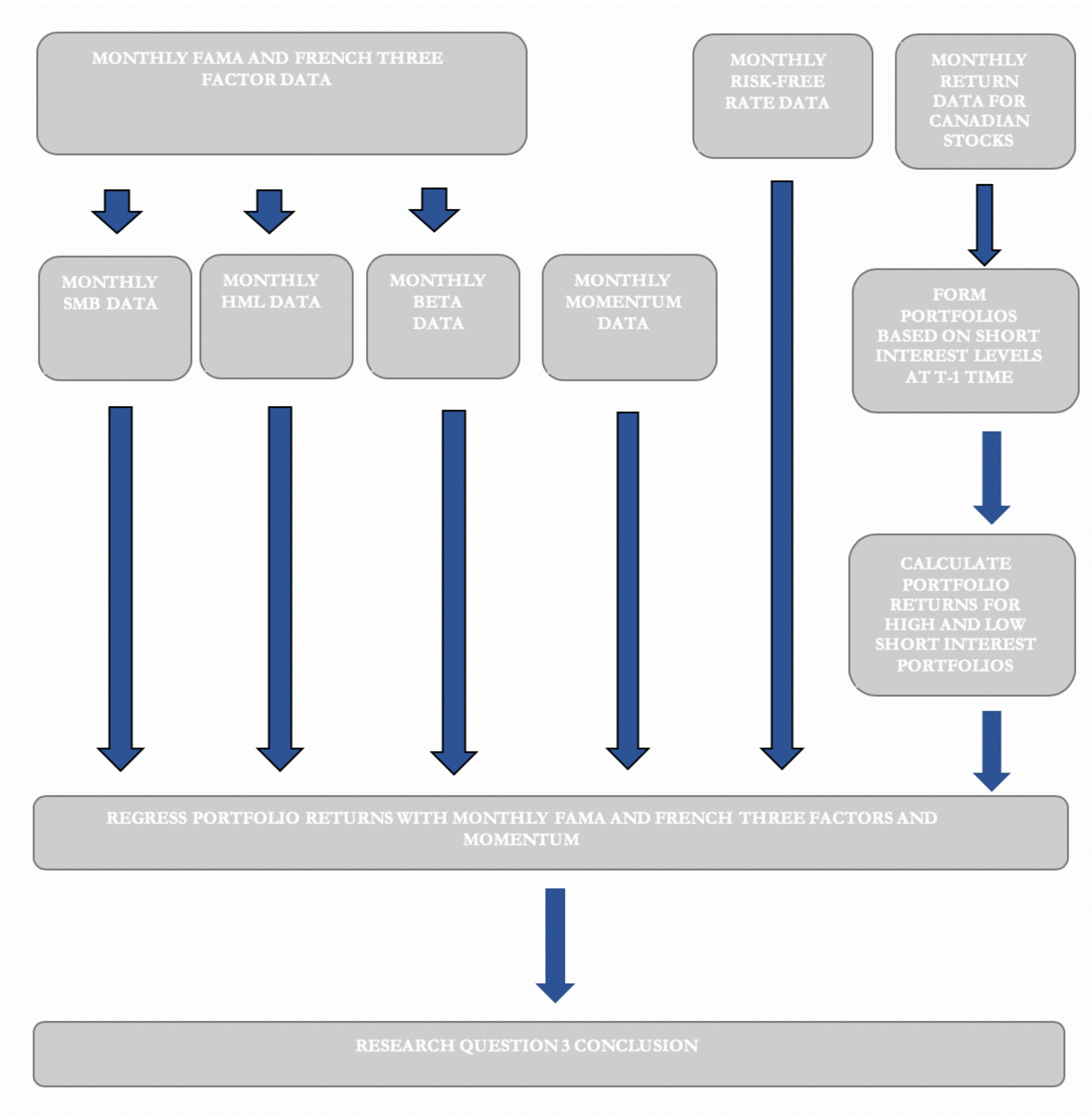
Figure A2: Research Question 2 Methodology Diagram



Notes: Figure A2 shows a methodology diagram outlining the steps that need to be taken in order to produce results for research question 2. The methodology diagram allows you to see clearly what data needs to be collected, what the data will help to produce and how from that I will achieve my conclusion for research question 2. Monthly Fama and French Three Factor data is collected, consisting of the factors of SMB (Small Minus Big), HML (High Minus Low) and Beta. Monthly momentum data is also collected. Monthly returns data is collected for US stocks and portfolios are formed on short interest levels at time t-1. Portfolio returns are calculated from this. This is combined with the monthly risk-free rate and excess portfolio returns are regressed against the three factors and momentum.

Appendix 4:

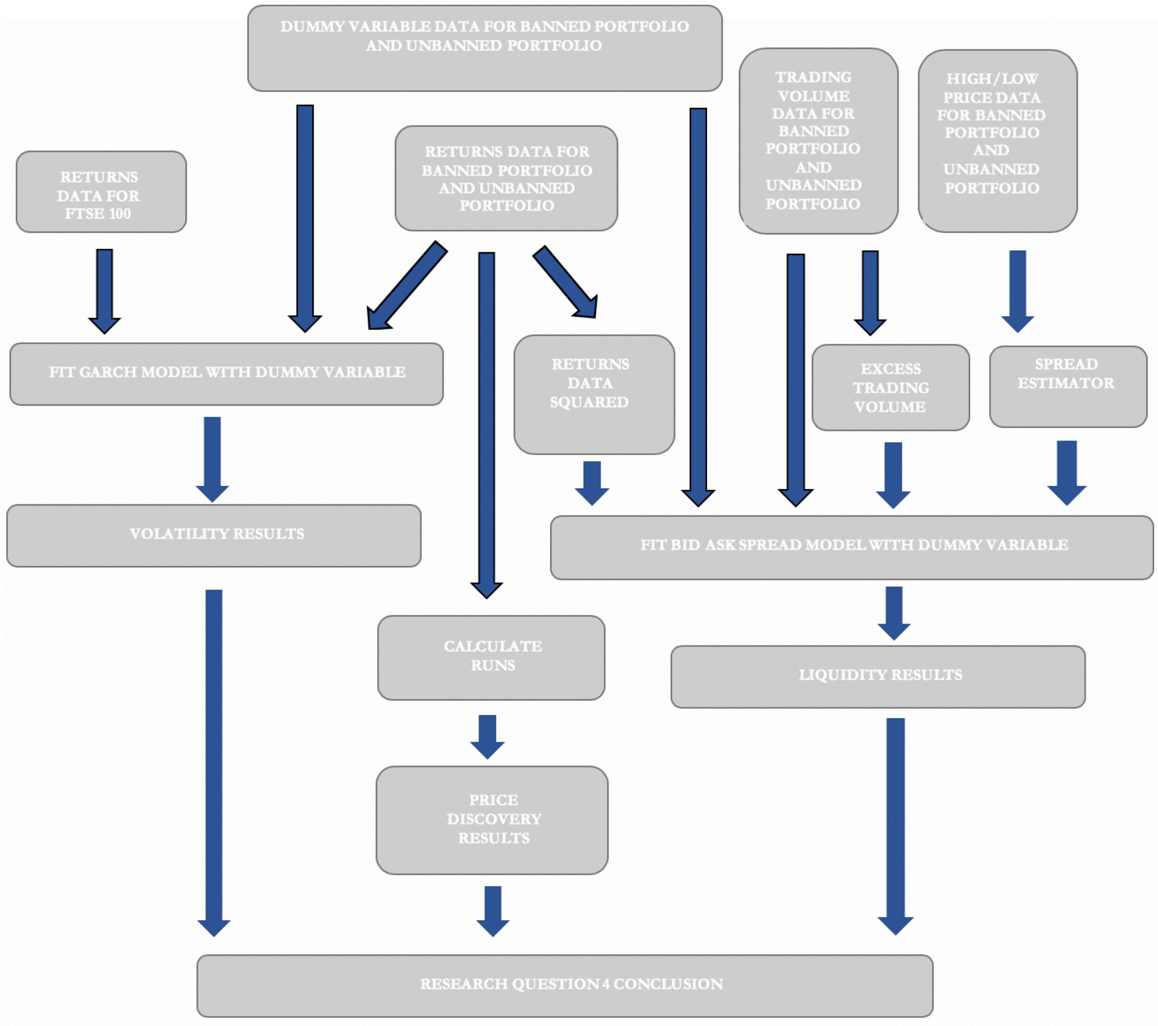
Figure A3: Research Question 3 Methodology Diagram



Notes: Figure A3 shows a methodology diagram outlining the steps that need to be taken in order to produce results for research question 3. The methodology diagram allows you to see clearly what data needs to be collected, what the data will help to produce and how from that I will achieve my conclusion for research question 3. Monthly Fama and French Three Factor data is collected, consisting of the factors of SMB (Small Minus Big), HML (High Minus Low) and Beta. Monthly momentum data is also collected. Monthly returns data is collected for Canadian stocks and portfolios are formed on short interest levels at time t-1. Portfolio returns are calculated from this. This is combined with the monthly risk-free rate and excess portfolio returns are regressed against the three factors and momentum.

Appendix 5:

Figure A4: Research Question 4 Methodology Diagram



Notes: Figure A4 shows a methodology diagram outlining the steps that need to be taken in order to produce results for research question 4. The methodology diagram allows you to see clearly what data needs to be collected, what the data will help to produce and how from that I will achieve my conclusion for research question 4. Returns data is collected for the FTSE 100, Dummy variable data, returns data, trading volume data and high/ low price data is collected for both banned and unbanned portfolios on a daily basis. The returns data for the FTSE 100, dummy variable data and returns data for the banned and unbanned portfolio is used to fit a GARCH model. This provides volatility results. The returns data for the banned and unbanned portfolio can also be used to calculate run distributions, which give the price discovery results. Squaring the returns data and using trading volume data for banned and unbanned portfolios, excess trading volume can be calculated. High/low price data for banned and unbanned portfolios can be used to obtain a spread estimator. The combination of the returns data squared, dummy variable data, trading volume data, excess trading volume data and the spread estimator can be used to fit a bid-ask spread model. This gives the liquidity results.

BLANK PAGE