Short Interest and the Cross-Section of S&P500 Share Returns

By

Kyle Cumming

CMMKYL001

A dissertation submitted to the Faculty of Commerce, University of Cape Town, in partial fulfilment of the requirements for the degree of Master of Commerce specialising in Finance in the field of Investment Management.

Abstract

This study aims to investigate the cross-sectional relationship between short interest and excess returns of the constituent securities of the Standard and Poor's 500 Index on a monthly basis. Short interest data is defined in relation to both trading volume and equity float for an 84-month period between January 2015 and December 2021 to examine the expected negative relationship. The use of the Fama-Macbeth (1973) method produces mixed empirical findings. The results of the short interest ratio do not support the findings of prior research, finding no significant relationship between the two variables. The short float ratio, however, produces a significantly positive relationship at the 10% level, supporting the "contrarian view". An increase in the short float ratio of 1% leads to a 19.6 basis point increase in excess return in the subsequent month. Overall, our results for the short interest ratio support the efficient market hypothesis. In contrast, the short float ratio serves as a bullish indicator.

Supervisor:

Professor Paul van Rensburg

October 11, 2022

Plagiarism Declaration

I, **Kyle Cumming**, hereby declare that the work on which this dissertation/thesis is based is my original work *(except where acknowledgements indicate otherwise)* and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university. I empower the University of Cape Town to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: .... Signed by candidate

Date: **October 11, 2022**

Acknowledgements

I would like to thank the following people, without whom I would not have been able to complete this research, and without whom I would not have completed my Master's Degree. My supervisor and course convener, Professor Paul van Rensburg, whose insight and knowledge into this topic have steered me in the right direction. His enthusiasm, encouragement, and patience is something I have greatly appreciated. A special thank you to my family and loved ones, for the unconditional support through both my coursework and dissertation. Your empathy, motivation, and love never goes unnoticed.

List of Acronyms

ADTV Average Daily Trading Volume

AMEX American Stock Exchange

BP Basis Point

CBOE Chicago Board Options Exchange

CHANGE_SFR Monthly Change in Short Float Ratio

CHANGE_SIR Monthly Change in Short Interest Ratio

ER Excess Return

ETF Exchange Traded Fund

EY Earnings Yield

FINRA Financial Industry Regulatory Authority

FTSE Financial Times Stock Exchange

LN_SFR Natural Logarithm of Short Float Ratio

LN_SIR Natural Logarithm of Short Interest Ratio

MOM Momentum Effect *(6 months)*

NASDAQ National Association of Securities Dealers Automated Quotations

NYSE New York Stock Exchange

PXREV Price Reversal Effect *(30 days)*

R Return

S&P500 Standard and Poor's 500 Index

SEC United States Securities and Exchange Commission

SFR Short Float Ratio

SI Short Interest

SIR Short Interest Ratio

TRI Total Return Index

VOC Dutch East India Company

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 The History of Short Selling

The origin of short selling dates back to 17th century Western Europe, where Isaac Le Maire, an ousted director of the Dutch East India Company (VOC), together with a consortium of Dutch businessmen, began shorting shares of the VOC on the Amsterdam Exchange in February 1609 (van Dillen, 1930). Le Maire and associates formed the *"Groote Companie"* which intended to profit from price declines of the VOC through the short selling of shares *"in blanco"*, where, at the time of the transaction, the seller was not in possession of the shares being shorted (Van Dillen et al., 2006). Furthermore, these shares were sold forward and promised future delivery in the next one to two year period (Bris et al., 2007). The Groote Companie has since been described as a bear-trading syndicate that sought to artificially deflate the VOC share price through the distribution of false adverse information (Van Dillen et al., 2006). In the following year, the bear raid, together with the nature of the short sales, resulted in the VOC share price declining by approximately 12%, while the profits of the Groote Companie, which held a net short position, mounted (Bris et al., 2007).

In January 1610, the directors of the VOC argued that the company was suffering speculative abuse and submitted a petition to the States of Holland that the controversial practice of short selling be banned, to which the Dutch government ultimately acceded, marking the first instance of government intervention in a securities exchange (De Marchi and Harrison, 1994). The ban was temporary and was often flouted by investors, however, despite continued opposition, the law was reinstated in 1624, 1630, 1636, and 1677, after which the Dutch government decided to instead impose taxes on the profits from short selling (Poitras, 2009).

Short selling has since become commonplace in modern financial markets and has grown in popularity in recent decades. In 1973, approximately 53% of short sales occurring on the New York Stock Exchange (NYSE) were conducted by exchange specialists attempting to maintain orderly markets, while only a negligible portion was held by speculators, with the remainder of short positions held by the public (Hanna, 1976). Brent et al., (1990) discovered that between 1977 and 1990, outstanding short interest on the NYSE increased by more than 500%, to the point where short positions accounted for approximately 9% of trading volume on the exchange in 1984. The authors stated that brokers and specialists momentarily utilized short selling to align their clients' supply and demand imbalances; hence, approximately 80% of short selling was conducted by specialists, albeit for a short period, to minimize their exposure to risk (Brent et al., 1990). Between 2000 and 2004, Boehmer et al., (2008) noted that short interest accounted for approximately 12.9% of total trading volume on the NYSE, with members of the public comprising between one and two percent of total shorting volume, whereas institutional shorts contributed about 74% and proprietary member firm shorts about 20%. Diether et al., (2009) found that during 2005, short-sellers contributed, on average, to 23.9% of share volume on the NYSE and 31.3% on the NASDAQ. According to Chang et al., (2019), one out of every five shares traded in US equity markets in 2017 was a short position, emphasising the prevalence of short sales in financial transactions.

1.2 Procedure of Short Selling

Short selling, also known as covered shorting, is the process in which a short-seller borrows a security from either an institutional investor or broker and sells the security on the secondary market prior to taking ownership (Smith 1968). Institutional investors, such as pension funds, index funds, and insurance companies, permit their assets to be lent out as their mandates often specify long-term investment horizons (Asquith et al., 2005). Duffie et al., (2002) therefore argued that through lending, institutions are afforded an opportunity to earn a risk-free return generated from the fees charged to short-sellers who have borrowed assets.

Luhr, (1992) stated that security lending agreements are a prerequisite to the finalisation of short sale transactions and are generally facilitated by brokers as opposed to individual investors. The lending agreement stipulates the conditions of the loan, such as the type of collateral, duration, interest rates, and the lender's fees (Luhr, 1992). Geczy et al., (2002) noted that during lending agreements, short-sellers would exchange a minimum of 102% of the market value for domestic securities and 105% of the market value for international securities, along with any accrued interest on debt securities, as cash collateral to the broker, acting in either their personal capacity or serving as an intermediary between short-sellers and institutional investors. Duffie et al., (2002) argued that the aforementioned percentages are dependent on the volatility of the asset being shorted, such that a more volatile asset requires additional collateral. While short-sellers are permitted to use other securities in lieu of cash collateral, Geczy et al., (2002) discovered that this occurred in only 1% of cases in the United States between 1998 and 1999. Jones and Lamont, (2002) further noted that these lending agreements generally occur on a continual or open basis, as renewal takes place daily with cash collateral adjustments dependent on the market price of a security *(mark-to-market)*.

Throughout the lending agreement, the lender pays the short-seller interest at an overnight rate, known as the rebate rate, on the cash collateral, which accrues daily and is settled at month-end (D'Avolio, 2002). The rebate rate serves as a method to establish an equilibrium between supply and demand in the lending market and, as a result, is renegotiated daily (Jones and Lamont, 2002). Reed, (2002) noted that large *(small)* market capitalisation stocks are usually inexpensive *(expensive)* to borrow because stock lenders are plentiful *(scarce)*, hence the rebate rate is high *(low)*. The rebate rate on stocks that are low-cost to borrow is known as the general collateral rate, which often closely tracks overnight market rates, such as the United States Federal Funds rate (D'Avolio, 2002). The cost of shorting, calculated as the difference between the general collateral rate and a comparable reinvestment rate, for these low-cost stocks is usually minimal (Jones and Lamont, 2002). Duffie et al., (2002) argued that a rebate rate below market rates is advantageous to the lender as it reduces funding costs, but disadvantageous to the short seller as higher interest could have been earned in the market.

Throughout the lending agreement, the title of ownership together with voting rights is transferred to the short-seller, with the short-seller also receiving the dividends from the borrowed security, which are then paid to the lender in cash (Duffie et al., 2002). A lender may issue a recall notice, where the short-seller is obligated to return the borrowed securities to the lender, hence terminating the lending agreement, with failure to comply resulting in the short-seller being referred to the regulatory authority, which may lead to sanctions and subsequent reputation risk (Geczy et al., 2002). Under normal conditions, the short-seller will often borrow shares from another lender and return them to the counterparty, thus covering the short position (Jones and Lamont, 2002). However, Angel and McCabe, (2009) argued that on occasion, both the short-seller and the broker are incapable of locating the required shares, requiring either party to purchase the shares outright. If the broker and short-seller fail to deliver the assets within the settlement period *(t + 3)*, the lender may purchase the shares outright at current market prices using the collateral posted in the lending agreement, with all additional costs borne by the short-seller (Reed, 2013). This risk of recall provides motive for a short-sellers preference towards institutional investors as their positions are often held for extended periods, whereas brokers' are less favoured, as their stock positions are more dynamic and recalls are likely to occur more frequently (Reed, 2013).

1.3 Short Selling Restrictions

Throughout history, the practice of short selling has been criticised by governments, particularly amid negative economic cycles (Meeker, 1932). De Marchi and Harrison, (1994) stated that short-sellers have been blamed for the majority of financial crises during the last 400 years, starting with the Dutch episode. However, Stanley, (2009) argued that speculators are often only betting into downward markets, not creating them.

During World War I, in November 1917, short selling regulations were enacted on the NYSE, after it had plummeted 31% in the previous year, with short-sellers being accused of demoralising the market (Meeker, 1932). Short-sellers were again condemned by President Herbert Hoover, along with many other politicians, investors, and journalists alike, who believed short-sellers were responsible for the Wall Street Crash of 1929 (Jones and Lamont, 2002). During the period of the Great Depression, multiple short selling regulations were introduced, such as the *"uptick rule"*, which required that the short sale must be greater than the last traded price of the security, and the Investment Company Act of 1940, which severely restricted mutual funds from shorting (Stanley, 2009).

Amid the global financial crisis, for the three weeks between September 2008 and October 2008, restrictions on short selling were imposed in the United States, as regulators feared that a downward market would be exacerbated through shorting (Fantazzini and Maggi, 2012). By examining intraday short selling data on the NYSE between 2005 and 2008, Bailey and Zheng, (2013) incorporated this financial crisis into their study, finding that while aggressive shorting and predatory trading may have had slight destabilisation effects, they were of little economic significance and typically disappeared within minutes. Therefore, Bailey and Zheng, (2013) found no justification for the short selling restrictions of 2008, arguing that restrictions resulted in delayed price discovery and lower liquidity while failing to prevent declines in stock prices.

During the Coronavirus pandemic in 2019, equity markets worldwide experienced an unprecedented decline while volatility surged. Enriques and Pagano, (2020) stated that the financial turmoil in European markets caused regulators across the continent to implement temporary short sale restrictions in Spain, Austria, Greece, Italy, Belgium, and France in an attempt to stabilise markets and regain investor confidence as speculative short sellers were again assumed to be asserting downward pressure on security prices. The empirical findings of Siciliano and Ventoruzzo, (2020), who investigated these major European stock exchanges, support those of Bailey and Zheng, (2013), as stocks that were banned from being sold short were found to be less liquid, have greater information asymmetry, and lower abnormal returns in comparison to their unbanned counterparts. Short selling restrictions are thus regarded as an inefficient instrument for market stabilisation during periods of high volatility (Bailey and Zheng, 2013; Siciliano and Ventoruzzo, 2020).

1.4 Naked Short Selling

Fotak et al., (2009) described *"naked short selling"* as a short-seller not arranging and/or intending to borrow an underlying security from a broker prior to shorting a stock, which has become an illegal practice in the United States and is facing increasing restrictions across financial markets. The situation in which short-sellers do not deliver the securities to the buyer within the settlement period is known as *"failure to deliver"* (Angel and McCabe, 2009). Failure to deliver can occur for legitimate reasons, such as routine failures due to transaction errors or long failures, in which the short-seller is unable to obtain deliverable shares because they are long in similar securities (Evans et al., 2009). Boni, (2006) listed the delayed conversion of convertible notes as another legitimate reason, however, the author further noted that in certain circumstances, the failure to deliver is intentional.

Autore et al., (2015) suggested that intentional failures occur when short-sellers consciously fail by not borrowing securities on the lending market. Evans et al., (2009) described that when short positions are anticipated to be covered quickly, short-sellers often fail to deliver as they avoid the hassle of locating and borrowing the required securities. Manipulative fails are another instance of intentional failures, with Angel and McCabe, (2009) arguing that when a security is sold short before being borrowed, short-sellers can potentially short sell infinite shares, thus reducing buyer interest and depressing stock prices to artificially low levels, making the strategy self-perpetuating and resulting in the shorted firm facing reputational damage and hampering their ability to obtain finance.

Regulatory authorities hence described naked short selling as a tool used in bear raids that disrupts financial markets by exacerbating price declines through momentum strategies (Khanna and Mathews, 2010). As a result, regulatory authorities in the United States, Britain, and many other countries across the globe placed restrictive regulations on naked short selling in the wake of the global financial crisis (Boulton and Braga-Alves, 2012). Interestingly, post-crisis research found that naked short selling was not a manipulative distorter of financial markets (Fotak et al., 2009). Instead, naked short sales were found to increase after credit downgrades, not before, indicating that naked short-sellers acted on public information instead of prompting price declines (Boulton and Braga-Alves, 2012). Regulatory decisions to restrict naked short selling are thus not supported by the majority of empirical evidence (Angel and McCabe, 2009; Fotak et al., 2009; Boulton and Braga-Alves, 2012).

1.5 The Benefits and Drawbacks of Short Selling

Although controversial, covered short selling has been found to have multiple positive market impacts. Diamond and Verrecchia, (1987) argued that shorting provides bearish investors an opportunity to participate in the market, which would not exist in long-only markets where short sales are restricted. Short-sellers are also found to actively drive stock prices towards their intrinsic value, hence improving pricing efficiency (Dechow et al., 2001). Desai et al., (2002) noted that short-sellers frequently identify firms that are experiencing financial difficulty and improve informational efficiency, as negative public information is impounded into stock prices more rapidly. In Hong Kong, Chang et al., (2007) found that short sales improved liquidity whilst simultaneously narrowing the bid-ask spread. Charoenrook and Daouk, (2009) examined short selling in 111 countries and found that when short selling is permitted, stock returns are, on average, significantly less volatile. Short selling has thus been found to improve market quality, depth, and efficiency during periods of market calm (Helmes et al., 2010).

However, regulators, practitioners, and academics have argued that short selling may have negative effects, particularly in extreme market downturns, by exerting downward pressure on already declining stock prices (Fotak et al., 2009). Angel and McCabe, (2009) suggested that short selling may incentivise illegal market conduct, as short-sellers are alleged to circulate misleading and false adverse information in an effort to depress a shorted firm's stock price. In contrast to prior literature, both Henry and Koski, (2010) and Shkilko et al., (2012) argued that short-sellers cause stock prices to diverge from their fundamental value and hence contribute to market instability. Aggarwal et al., (2010) stated that short selling causes a decoupling between ownership and voting rights, where short sellers may vote for unfavourable decisions to turn a profit, hence, institutional investors often issue a recall notice for borrowed stock prior to material proposals. Although the empirical evidence of the aforementioned is sparse, the fears associated with short selling have resulted in a multitude of rules and regulations being implemented by regulatory authorities worldwide (Schindler, 2015).

1.6 Research Question

This study aims to determine whether short interest has an effect on the cross-section of share returns of the constituent securities of the Standard and Poor's 500 Index (S&P500 Index) over an 84-month period between January 2015 and December 2021?

This period has been chosen since it consists of the most up-to-date data available. While many European authorities enforced temporary short selling bans during the pandemic, the United States Securities and Exchange Commission (SEC) did not enact any restrictions (Enriques and Pagano, 2020; Siciliano and Ventoruzzo, 2020). In accordance with Schindler, (2015), the reliability and validity of the empirical results in this study will not be deteriorated due to the absence of significant legal regulations in the sample.

1.7 Outline of Paper

The remainder of this paper will be organised as follows. Section 2 contains the literature review, which will discuss the motives for selling short, along with the three distinct relationships that have been reported in prior research, whilst also suggesting the psychological reasoning for short selling, together with a summary of the past empirical findings. In Section 3, the data is described. Section 4 discusses the methodology which will be used for the cross-sectional approach while also developing the hypothesis to be tested. Section 5 both presents and discusses the descriptive statistics of the data. In Section 6, the Fama and MacBeth (1973) regression approach is introduced. Section 7 reports the empirical findings of the study, while Section 8 tests for robustness. Section 9 analyses the cumulative returns to both short interest measures over the sample period, with Section 10 outlining the recommendations for further research. Finally, Section 11 comprises of concluding remarks.

2 Literature Review

2.1 Motives for Short Selling

Prior research has suggested numerous reasons for market participants short selling a stock. According to Dyl, (1978), investors may short a security for either taxation minimisation or taxation avoidance purposes. This strategy is known as *"shorting against the box"* and occurs when an investor simultaneously takes a long position while holding an equivalent short position on the margin account (Brent et al., 1990). A neutral position is therefore created in an attempt to generate a profit that will only be recognised in a later period at a more favourable taxation rate, such that any taxable gains are deferred (Arnold et al., 2005). Dyl, (1978) further argued that the benefit of delaying taxable gains must exceed the opportunity cost of not receiving the proceeds from the immediate sale, in addition to any interest earned had the proceeds been invested in a short-term liquid asset. According to Oesterle, (2006), investors who use the box strategy are frequently unwilling to close out their long position because the asset is expected to increase in value over time. Brent et al., (1990) listed various factors to be considered prior to shorting against the box, such as the current and future taxation rate, the prevailing interest rate, the investor's time horizon, and the magnitude of capital gains.

Renshaw, (1977) proposed that short selling could also be used for arbitrage and hedging purposes. Since a perfect hedge is incapable of yielding any profits, an additional financial instrument, such as index futures, stock options, and convertible bonds, is utilised to ensure that the value between the underlying security and the short position is correlated, which therefore allows arbitrage profits to be realised (Ramsay, 1993). Werner, (2010) argued that shorting stock for the following motive has gained popularity in recent decades as a result of the explosive growth of hedge funds, particularly in the United States, that employ market-neutral strategies. The author further argued that these uninformed short-sellers have reduced the accuracy of short interest as a measure of negative investor sentiment (Werner, 2010).

Short selling is most notably motivated by speculation, where bearish investors attempt to profit off an overpriced security through a declining stock price (Baron and McDonald, 1973). Ramsay, (1993) categorised speculative short-sellers as information-based traders whose beliefs are contrary to the market consensus. Brent et al., (1990) further stated that securities with a lower *(higher)* variance in returns are likely to have fewer *(greater)* speculative short-sellers, implying that residual variance provides an indication of the heterogeneity of investor opinion. By short selling a security whose market price has deviated from its intrinsic value, short-sellers aid in the price-determination process while also improving market efficiency (Woolridge and Dickson, 1994). Staley, (1996) stated that through speculative shorting, liquidity is provided to the market by creating supply when large orders result in trade imbalances and volatile prices. In contrast, Haruvy and Noussair, (2006) found that speculative short sellers can potentially create supply and demand imbalances, resulting in a divergence between market prices and fundamental values.

2.2 Different Relationships found in Prior Research

While market participants often have different incentives for short selling a security, there are conflicting opinions on the relationship between the level of short interest and a security's future price performance. Researchers have been unable to reach a consensus due to conflicting empirical results; hence, three distinct categories of the relationship have emerged.

2.2.1 Evidence of No Relationship - *Nonpredictive*

The first category finds that no statistically significant relationship exists between the level of short interest and future stock returns. The first advocate for this line of reasoning was Mayor, (1968), who investigated the S&P500 Index between 1962 and 1966 by utilising a multiple regression analysis with alternative lags and a Monte Carlo simulation technique, with the results of each reporting no statistically significant relationship between the two variables in the cross-section. Instead, Mayor, (1968) provided evidence supporting the *"random walk hypothesis"*, which states that unless trends are present between the two variables, no significant correlation would exist between the prior short interest level and the stock price in the following period. Therefore, an investor analysing short interest levels would be incapable of predicting the price movements of an equity in the next period and would perform no better than under a chance model (Smidt, 1968). Pinches, (1970) similarly found that no discernible relationship was present, however, the author further argued that if a correlation were to exist between the two variables, investors would be able to consistently beat the market and would presumably exploit arbitrage opportunities, which over time would eliminate both future profits and the predictive power of short interest.

Smith, (1968) investigated companies on the American Stock Exchange (AMEX) between 1967 and 1968, reasoning that since the sample had fewer shares outstanding and was less actively traded, the effect of short interest on subsequent return was expected to be more significant than that of the NYSE, which was a deeper, less volatile market. By using average volume traded as a proxy for the shares in supply, Smith, (1968) found that the price impact of short interest was more substantial for securities with a lower supply of shares. This finding is supported by Ackert and Athanassakos, (2005) and Schindler, (2015) whom both suggested including a control variable for firm size. In addition to reporting no relationship, stocks with higher short positions were found to be more price volatile, increasing *(decreasing)* more rapidly in market upswings *(downswings)*, illustrating that short interest has informational value (Smith 1968).

Hurtado-Sanchez, (1978) further noted that the level of short interest neither impacts the current nor future stock price. Instead, short selling was found to be a mechanism that aligned the risks and returns of a security, thus reducing excess returns and aiding in market stabilisation

(Hurtado-Sanchez, 1978). Caster and Vu, (1987) expected that shares held short for speculative purposes would experience a decline in price following the announcement of the short interest level. However, the author's empirical findings suggest that short selling increased together with rising stock prices in the pre-announcement period, yet post-announcement the decline in stock prices, on average, was insignificant. Therefore, according to Caster and Vu, (1987), an increase in short interest serves as neither a bearish nor a bullish indicator. Brent et al., (1990) support this view, confirming that investors who base their trading strategies on changes in short interest levels would be inefficient because the effect on stock price, especially in the short run, is largely insignificant. Similarly, Woolridge and Dickson, (1994) argued that on a month-to-month basis, the level of short interest does not result in investors earning abnormally high or low returns. Hence, short interest fails to serve as a reliable predictor of return in the next period.

Although not directly related to this study, Mayor, (1968), Brent et al., (1990), and Woolridge and Dickson, (1994) concurrently investigated whether a relationship existed between short interest and stock returns in the time series. However, the empirical findings are synonymous with those reported in the cross-section analyses, indicating that no relationship is present between the two variables. The perception that short sellers drive stock prices down to artificially low levels through short selling to earn an excess profit at the detriment of less informed market participants is seen to be false (Woolridge and Dickson, 1994). Instead, short-sellers are found to provide liquidity by shorting during upward market movements and covering during downward market movements (Staley, 1996).

Through demonstrating the irrelevance of short interest as a predictor of subsequent stock return, the findings of Mayor, (1968) were consistent with the *"efficient market hypothesis"*. Pinches, (1970) stated that a characteristic of efficient markets is that the market price of an asset, at any time, is expected to closely represent the asset's intrinsic value. Therefore, both the positive and negative implications of large short interest levels should be inherent in an asset's current trading price (Caster and Vu, 1987). Both Hurtado-Sanchez, (1978) and Figlewski, (1981) argued that short selling is the primary method through which price expectations from capital asset pricing models are realised by aligning returns to be commensurate with risk, and that the disparity between expected and realised returns is primarily a result of the regulations surrounding short sales in the market.

Brent et al., (1990) further found that stocks that had high betas, convertible securities, and options had more shares shorted, implying shorting was used for arbitrage and hedging purposes, particularly during volatile market periods. In addition, the authors found a slight seasonality effect, as short positions increased towards year-end in an attempt to defer capital gains and losses for taxation purposes (Brent et al., 1990). Asquith et al., (2005) aimed to differentiate short-sellers depending on their motive, as pessimistic short-sellers took positions based on firm valuation and acted as informed traders who impounded negative information into stock prices. In contrast, short sellers who took positions driven by arbitrage and hedging strategies were characterised as uninformed short sellers (Asquith et al., 2005). Desai et al., (2002) attempted to distinguish between valuation and arbitrage shorts ex-ante, noting that firms categorised as valuation shorts displayed high levels of short interest as well as poor subsequent returns, whereas firms categorised as arbitrage shorts did not display significantly negative future returns. Recent literature has found evidence of an increase in short selling caused by arbitrage and hedging activities due to the market-neutral strategies which multiple hedge funds employ, which has resulted in the negative relationship between short interest and future return weakening substantially (Werner, 2010). As a result, Werner, (2010) argued that informational content of short interest has declined, along with its precision as a measure of negative investor sentiment.

2.2.2 Evidence of a Positive Relationship - *Bullish Indicator*

The second category finds that a statistically significant positive relationship is present between the level of short interest and the subsequent price performance of an equity. This contrarian view is widely shared by financial analysts and technical traders and has occasionally been referred to as *"Wall Street Wisdom"* (Epstein, 1995). Both Seneca, (1967) and Mayor, (1968) argued that at a certain point, each short position will need to be covered, resulting in future share purchases that will place upward price pressure on shorted assets. A high level of short interest is therefore seen to reflect a latent demand for stocks that are sold short (Smidt, 1968).

Due to the majority of institutional investors being restricted from selling short, high levels of short interest are also seen to be indicative of potentially unwarranted pessimism from traders and the general public (Biggs, 1966).

Hanna, (1976) investigated whether increases *(decreases)* in the short interest ratio were related to investor pessimism *(optimism)* towards overvalued *(undervalued)* stocks. The empirical findings suggest that investors tend to systematically over-discount events even though the vast majority have similar levels of optimism *(pessimism)* surrounding the future share price (Hanna, 1976). The components of the short interest ratio, being short interest *(numerator)* and average daily trading volume *(denominator)*, prove to be unsuccessful predictors in isolation. However, as a whole, increases *(decreases)* in the short interest ratio act as a bullish *(bearish)* indicator in the cross-section (Hanna, 1976). Consequently, an investor purchasing a stock with a high short interest ratio would outperform a chance model (Hanna, 1976). The remaining literature on the positive cross-sectional relationship is sparse; thus, empirical findings from the time series are also discussed for completeness. The findings of Hanna, (1976) are partially supported by Aksu and Gunay, (1995) who tested for co-integration between short interest levels, average daily trading volume, and market price. Although the authors found no evidence of a positive or negative relationship, their findings suggest that the variables are contemporaneously interrelated in a positive and statistically significant way (Aksu and Gunay, 1995).

Schindler, (2015) used monthly data to investigate the NASDAQ-100 Index between 2012 and 2014. The authors' empirical findings corroborate those of Hanna, (1976), noting a statistically significant positive relationship exists between short interest and subsequent stock return. Furthermore, short interest, whether defined in relation to equity float or average daily trading volume, produces the same positive relationship, however, the relationship is slightly stronger for the latter divisor (Schindler, 2015). Consistent with Ackert and Athanassakos, (2005), Schindler, (2015) further found that a positive relationship was also evident between firm size and abnormal return.

Mohamad et al., (2016) investigated the effect of short interest on the price of Exchange Traded Funds *(ETFs)* on the London Stock Exchange between 2006 and 2010. By construction, an

ETF is designed to track a basket of assets, therefore, by investing in these securities, market participants gain exposure to a variety of stocks. Hence, ETFs are shorted by both speculators and hedgers (Huang et al., 2021). Mohamad et al., (2016) found that ETF short positions are primarily taken by bullish hedgers in an attempt to protect their portfolios, and as a result, when ETFs are heavily shorted, there is a positive abnormal return in the subsequent period. However, the nature of an ETF makes the informational content of short interest difficult to interpret in this context (Mohamad et al., 2016; Huang et al., 2021).

Fosback, (1993) stated that to a contrarian investor, extreme levels of short interest signal that market participants have become excessively bearish and that an upward price reversal is likely to occur. This is supported by Au et al., (2009), who stated that at the extremities, short interest transforms from a bearish to a bullish indicator, such that excessive short interest levels precede positive future returns. In these instances, the potential for a *"short squeeze"* to occur is heightened (Schindler, 2015). Vryghem, (2017) argued that short squeezes can occur when a stock price unexpectedly rises due to favourable news, such as operational announcements surrounding debt restructuring or approved patents; better than anticipated earnings announcements; increased trading activity regarding incoming buy orders; index rebalancing and share buybacks, as well as announcements regarding acquisitions or mergers. This favourable news acts as a catalyst for price increases, as short-sellers try to minimise their losses by covering their short positions (Cristian and Raisa, 2018).

In recent times, social media platforms such as Reddit and Twitter have also contributed to short squeezes occurring in heavily shorted stocks such as GameStop and Tesla, as retail investors pool their finances to sabotage short positions held by hedge funds, while simultaneously emphasising the fundamental vulnerability of financial systems (Chohan, 2021). The coincidental covering of short positions during a squeeze leads to increased share prices, attracting buyers who further intensify price spikes, resulting in a type of prisoner's dilemma occurring on a macro-economic scale (Vryghem, 2017).

2.2.3 Evidence of a Negative Relationship - *Bearish Indicator*

The third category, or *"speculative view"*, suggests that a negative relationship exists between the level of short interest and subsequent stock returns and is a commonly accepted stance amongst academics, particularly in modern research. The first advocate for this line of reasoning was Seneca, (1967), who conducted a regression analysis on the S&P500 Index and the outstanding short interest level between 1946 and 1965 using a fifteen-day lag, which is attributable to short interest data only being published by the exchange mid-month. Seneca, (1967) argued that although short positions are a commitment to purchase securities at a later date, the fundamental judgement of a short sale is an expectation of prices to fall in the near future. The empirical findings suggest a negative relationship exists between the two variables, implying that short interest serves as a bearish indicator (Seneca, 1967). Figlewski, (1981) also concluded that a negative relationship exists between short interest and excess return on the S&P500 between 1973 and 1979. Furthermore, the restrictions on short sales were found to be of particular importance, as they have an asymmetrical effect on investors who have either favourable or unfavourable firm-specific information (Figlewski, 1981). When regulations are present, optimistic return forecasts receive a disproportionately greater weighting than pessimistic forecasts, resulting in informational asymmetry in the market (Figlewski, 1981).

Diamond and Verrecchia, (1987) aimed to determine the impact that short selling constraints would have on the rate of adjustment of asset prices to new information, finding that an unexpected increase in short positions signals adverse news to market participants, hence causing a decrease in price, implying a negative relationship exists between short interest and future return. In addition, when short selling restrictions are present in the market, there is a notable rise in options trading as bearish investors write calls and purchase puts in an attempt to circumvent regulations whilst indirectly incorporating unfavourable information into the underlying stock price (Diamond and Verrecchia, 1987). Aitken et al., (1998) later examined the Australian Securities Exchange on an intraday basis between 1994 and 1996, where short interest levels were reported instantaneously, hence providing greater market transparency. Using a calendar-time portfolio approach, Aitken et al., (1998) findings suggest that when no lag is present in the reporting of short positions, short interest levels are viewed as a significantly bearish indicator and that abnormal negative returns follow soon after.

Angel et al., (2003) investigated the NASDAQ and noted that excessive short selling was found to typically precede negative stock returns. In addition, short selling activity seemed to occur rather consistently throughout a trading week, indicating that no day-of-the-week effect is apparent (Angel et al., 2003). Ackert and Athanassakos, (2005) studied short selling on the Canadian Securities Exchange, with the relationship between short interest, defined in relation to average daily trading volume, and excess returns found to be contemporaneously negative. Boehmer et al., (2008) analysed daily data to determine the effect of short positions on stock prices on the NYSE between 2000 and 2004. By utilising the Fama and MacBeth (1973) method as a regression approach, stocks that were heavily shorted seemed to underperform, on average, relevant to lightly shorted stocks in the cross-section, providing additional evidence of a negative relationship (Boehmer et al., 2008).

Similarly, Au et al., (2009) utilised daily data together with the Fama and MacBeth (1973) regression method to investigate short selling on the Financial Times Stock Exchange (FTSE) 350 Index in the United Kingdom between 2003 and 2006. The authors found evidence of a negative relationship between the level of short interest and abnormal stock returns, particularly amongst stocks with a high level of idiosyncratic risk (Au et al., 2009). Hence, stocks with lower *(higher)* levels of idiosyncratic risk were more *(less)* heavily shorted, thus, idiosyncratic risk is found to be a better deterrent to short selling than the associated transaction costs (Au et al., 2009). Akbas et al., (2013) noted, using the same approach, that the level of short interest is able to predict, not cause, abnormal negative stock returns in the subsequent period. Callen and Fang, (2015) also utilised the Fama and MacBeth (1973) method to investigate the impact of short interest on public United States firms between 1981 and 2011, and found that a high level of short interest serves as a predictor of stock price crash risk one-year ahead, hence providing further evidence of the negative relationship between short interest and subsequent return.

Zhu et al., (2019) studied the effect of trends in short selling on the cross-section of return on the AMEX, NASDAQ, and NYSE between 1988 and 2014, with trends in short positions proxied by long-term changes in the level of short interest. The trend in short interest seemed to provide additional predictive information beyond the current short interest level, suggesting that stocks with a substantial upward *(downward)* trend in short interest levels display significantly negative *(positive)* abnormal returns (Zhu et al., 2019). Nezafat et al., (2019) then investigated the relationship between hedge fund holdings and short interest on the same exchanges over a similar period, utilising the level of short interest and aggregated long positions as proxies to measure the opinions of different market participants, with disagreements between investors examined to determine the impact on subsequent returns. Stocks with high levels of short interest and large hedge fund holdings were found to have no abnormal returns, whereas stocks with high short interest and low hedge fund holdings exhibited abnormal negative returns in the cross-section (Nezafat et al., 2019).

Although not the focus of this study, prior research has also investigated the time series, which is briefly discussed, however, the majority of findings tend to correspond with the relationship found in the cross-section. Kerrigan, (1974) discovered that aggregate short interest is useful in forecasting future S&P500 returns, with a negative relationship reported. However, the author further suggested that short interest is a more effective predictor when the forthcoming market direction is known. The number of shares shorted is also seen to be of minimal importance, instead, Kerrigan, (1974) argued that the short interest ratio's apparent value lies in the denominator.

By utilising the Fama and French, (1993) four-factor model to examine the AMEX, NYSE, and NASDAQ, Boehmer et al., (2010) found further evidence of a negative relationship, stating that heavily shorted stocks significantly underperformed, regardless of whether short interest is defined in terms of the aggregate number of shares shorted, the aggregate short interest ratio or the median days-to-cover ratio. A multitude of authors argue that while short interest can potentially serve as an indicator of future return, short interest's informational content can be enhanced when examined concurrently with other variables (Au et al., 2009; Callen and Fang, 2015; Nezafat et al., 2019; Zhu et al., 2019).

2.3 Findings Associated with Short Interest

Whilst the recent amongst academics denotes a negative relationship between short interest levels and subsequent return, prior empirical research has indicated that numerous variables are associated with short interest. Desai et al., (2002) suggest that short-sellers may have access to private information and that short-sellers are found to select firms with high liquidity and a mismatch between market price and fundamentals. Boehmer et al., (2008) further stated that short-sellers are exceptionally well informed, with institutional short-sellers having both recognised and acted on crucial value-relevant information which is not yet impounded into a firm's stock price. Akbas et al., (2013) examined the informational advantage that short-sellers supposedly possess, arguing that this may either arise from a short-sellers ability to process and trade on information more efficiently than other market participants or that trades may be based on informational leaks. Callen and Fang, (2015) supported this argument by stating that the source of a short-sellers informational advantage is often questionable, as short positions may be based either on superior information analysis or private information.

Engelberg et al., (2012) investigated short-sellers on the NYSE between 2005 and 2007 and the effect which public news events obtained from the Dow Jones and Wall Street Journal had on short interest and subsequent stock returns over the period. Engelberg et al., (2012) argued that short sellers were neither uncovering nor anticipating news surrounding ratings, earnings projections, or joint ventures, however, there was a notable increase in short trading following news announcements, indicating that short sellers, like their peers, trade on publicly available information. This is supported by Lasser et al., (2010) who stated that the information which short-sellers utilise is exogenous and thus cannot be manipulated. Akbas et al., (2013) later found contradictory evidence, stating that short-sellers can predict changes in firm fundamentals several months prior to the information becoming public, with the ability to predict unfavourable earnings, forecast downgrades, and anticipate adverse company-specific news correctly being the predominant reason for short-sellers accuracy in predicting future negative returns.

The empirical evidence of Engelberg et al., (2012) implies that while short-sellers do not alter the market's perception of value, their advantage over other traders is primarily a result of their superior public information processing ability, which allows short-sellers to more accurately anticipate subsequent negative returns. Diamond and Verrecchia, (1987) argued that by reducing short selling costs, negative information will be incorporated into the market more rapidly, which will result in the excess return distribution being less negatively skewed. As a result, Aitken et al., (1998) continued to advocate for short sale transparency so that the price discovery process is more efficient. Zhu et al., (2019) found that while short-sellers are sophisticated and well-informed investors, the market tends to react slowly to the signal which trends in short interest portray.

Dechow et al., (2001) suggest that short-sellers frequently position themselves in securities that are forecasted to have lower subsequent returns, basing their stock-picking decisions on firms which have low fundamental-to-price ratios, such as price-to-book, price-to-earnings, cash-flow-to-price, and value-to-market ratios. Dechow et al., (2001) further argued that short-sellers utilised fundamentals to determine overpricing and that when these ratios mean reverted, short-sellers were found to cover their positions, which realigned the stock price with its intrinsic value. These empirical findings are supported by Desai et al., (2002), who stated that the aforementioned ratios are used by short-sellers in the stock selection process. Short-sellers are therefore seen to add liquidity to rising markets through shorting (Albert et al., 1997).

According to Boehmer et al., (2010), short-sellers are found to be partially successful in selecting overpriced stocks to short, yet they are equally successful in the recognition of underpriced stocks. Callen and Fang, (2015) indicated that short-sellers are able to detect negative information which company managers may be hoarding for equity incentives, accrual manipulation, or taxation avoidance purposes, whilst further being able to identify firms with substantial information asymmetry between company management and shareholders, weak governance mechanisms, and a tendency to take on excessive risk. Angel et al., (2003) thus described the investment approach of short-sellers as contrarian, as short-sellers tend to favour shares with increasing prices as there is an expectation that a mean reversion will occur and present profit opportunities. Ackert and Athanassakos, (2005) suggested that stocks with options and convertible bonds are found to extenuate the negative relationship between short interest and excess returns since both provide an alternative method of establishing a short position, often at a lower cost. Phillips, (2011) further stated that options allow investors an opportunity to impound negative information into stock prices when short-sale constraints are apparent, thus improving informational efficiency on stock exchanges.

Diamond and Verrecchia, (1987) noted that the borrowing costs for short selling were found to be largest when the stock valuation was high and subsequent returns were predicted to be low, as this is another characteristic of overpricing. Short-sellers were also found to refine their positions to assets with lower transaction costs in order to maximise their returns, whilst distinguishing whether low fundamental-to-price ratios were a result of momentary price surges, or due to temporarily low fundamentals (Dechow et al., 2001). Ackert and Athanassakos, (2005) findings suggest that investors short sell dual-listed stocks in Canada as opposed to America to benefit from lower transaction costs, implying that lower costs and fewer restrictions surrounding short sales will aid in improving market efficiency. Asquith et al., (2005) therefore concluded that the profitability of short positions is dependent on both the transaction and implementation costs associated with short selling.

Albert et al., (1997) investigated the effect of short interest on the NASDAQ over a period that did not prohibit naked short selling. In addition, during the sample period, the exchange did not conform to the uptick rule, which sought to reduce short-sellers exacerbating price volatility by ensuring that short sale prices were greater than the last traded price (Albert et al., 1997). Albert et al., (1997) found that short-sellers were earning excess returns on the NASDAQ, suggesting a negative relationship was present between short interest and future return. However, the authors failed to find evidence of market destabilisation, hence refuting the claim that short-sellers sell into falling markets and worsen the decline. Following contrarian investment schemes, short-sellers assume that increased volatility is partly due to investor overreactions, with Angel et al., (2003) proposing that volatile stocks allow for the associated costs of short selling to be covered and profit to be generated, as lower volatility stocks often fail to cover both transaction and implementation costs. The authors further found that short selling varied based on the level of stock price volatility, with more volatile assets being more heavily shorted and declining monotonically as volatility decreased (Angel et al., 2003).

By utilising a size variable, Boehme et al., (2006) noted that small market capitalisation stocks with high short interest levels frequently underperformed relative to the market. This is consistent with the findings of Asquith et al., (2005), who noted that smaller capitalisation stocks produced abnormally negative returns when high short interest levels were present. Asquith et al., (2005) further noted that the magnitude of returns from short selling was also dependent on institutional ownership, such that the lower the level of institutional ownership, the more negative the abnormal returns. Additionally, Akbas et al., (2013) found that short-sellers were better informed about assets with low institutional ownership, which often prove more difficult to short. Ackert and Athanassakos, (2005) found further evidence that size is an important determinant in the magnitude of abnormal returns on the Canadian Securities Exchange, with a positive relationship being reported between the two variables. Ackert and Athanassakos, (2005) reasoned that since smaller firms have a constrained supply of shortable shares, their excess returns are more negative.

Boehmer et al., (2010) found that stocks with a high share turnover and a low level of short interest are both economically and statistically undervalued and produce positive abnormal returns, which are greater, in absolute terms, than the negative returns of stocks that are heavily shorted. Angel et al., (2003) further argued that short-sellers were found to prefer higher-volume stocks, as the probability of a short squeeze occurring is lessened due to substantial liquidity, as investors are able to cover their positions more easily when stock prices experience an upward run. The empirical findings of Dechow et al., (2001) implied that short interest levels, combined with additional firm-specific information, served as an indicator of the subsequent stock return, with stock prices declining monotonically in relation to the level of short interest apparent.

Au et al., (2009) found contradictory evidence, stating that the negative relationship between short interest and subsequent returns is non-monotonic, thus suggesting that the use of short interest as a linear investment signal would be suboptimal. Instead, Au et al., (2009) argued that short interest was found to have diminishing returns, initially acting as a bearish indicator before reaching a maximum, where extreme levels of short interest then served as a bullish indicator due to an increased likelihood of a short squeeze occurring.

2.4 Short Selling Psychology

Platt, (2004) investigated the psychological reasoning that motivates investors to short sell. The author recognised that the demand for short selling is predominantly motivated by two investor emotions, namely fear and greed. Platt, (2004) further argued that the demand to short sell based on fear arises as companies in financial distress near and occasionally file bankruptcy, which is supported by the findings of Desai et al., (2002). Short selling a stock based on investor fear is found to be inversely related to a share's market price, such that a distressed firm with a low stock price is a more desirable fear-based short (Platt, 2004). When the demand to short sell is centered on investor greed, short-sellers often believe that a security is overpriced, thus, a positive relationship is present between high market prices and the greed-based demand for short selling (Platt, 2004). Therefore, Platt, (2004) stated that although the total demand for short selling is constrained by the supply of shortable shares, short positions are predominantly motivated by the previously mentioned psychological explanatory variables.

2.5 Summary of Empirical Findings

The empirical findings of prior research have utilised different methodological approaches, ranging from multiple regression analyses to the Fama and MacBeth (1973) method, calendar-time portfolio approaches, and the Fama and French, (1993) method, among others. The results of previous studies are inconclusive in their findings concerning both the magnitude and relationship present between short interest and subsequent stock returns. Three main categories surrounding the relationship between the two variables have developed, which are discussed above.

Research prior to the 1990's generally found that no statistically significant relationship was present. Mayor, (1968) used a multiple regression analysis to investigate the relationship, regressing closing prices on short interest levels that were non-lagged as well as lagged by one week, two weeks, three weeks, one month, three months, and, six months, with the results of all specifications reporting no evidence of a relationship between short interest and future return. Although supported by market participants and technical traders, the contrarian view, which states that a positive relationship is present between short interest and subsequent return, has received little empirical support (Hanna, 1976; Aksu and Gunay, 1995; Schindler, 2015).

The most widely dispersed view amongst academics, particularly in modern research, is that a negative relationship exists between short interest and future stock returns, which implies short-sellers are speculators who attempt to profit from a declining stock price. Callen and Fang,

(2015) utilised the Fama and MacBeth (1973) method, setting the lags to three months, six months, twelve months, and two years. The results indicate that short interest can predict stock price crash risk up to one year ahead, further emphasising the presence of a negative relationship (Callen and Fang, 2015). Zhu et al., (2019) employed the same regression approach with no lag and concluded that upward trends in the short interest level result in statistically significant and negative excess returns. Desai et al., (2002) concluded that the negative relationship persists for a period of up to twelve months and further argued that firms that are significantly shorted have a greater probability of delisting and/or liquidation in the subsequent 36 months after substantial short interest levels are reported.

In contrast, Asquith et al., (2005) found that highly shorted assets underperform for brief periods, thus, for negative abnormal returns to be realised frequent portfolio rebalancing, which proves costly, is necessary. Boehmer et al., (2008) later noted that the decline in prices due to short selling is permanent, suggesting that short sellers are neither manipulating nor temporarily deflating stock prices, but contributing towards market efficiency. The conflicting relationships reported between short interest and future returns thus warrant further research. The reasons for which empirical findings may differ likely arise due to the market and time period examined, the methodology and model specification utilised, and the chosen time lag.

Prior empirical research is distributed across a range of markets, as Aitken et al., (1998) investigated the effect of short interest in Australia, while Ackert and Athanassakos, (2005) examined the relationship in Canada, and both Au et al., (2009) and Mohamad et al., (2016) studied the effect of short interest in the United Kingdom. However, the vast majority of prior research is centered on the American market (Seneca, 1967; Kerrigan, 1974; Figlewski, 1981; Angel et al., 2003; Boehmer et al., 2008; Callen and Fang, 2015; Schindler, 2015; Nezafat et al., 2019; Zhu et al., 2019). Due to data availability and the abundant comparable literature, this paper will investigate the effect of short interest and the cross-section of S&P500 share returns.

3 Data

S&P Dow Jones Indices manages the S&P500 Index, which is comprised of the 500 largest publicly traded companies on the Chicago Board of Options Exchange (CBOE), NASDAQ, and NYSE, weighted by float-adjusted market capitalisation (S&P Dow Jones Indices, 2022). Float-adjusted indicates that only shares that are traded by the public are considered, while shares held by other companies, the government, and management are excluded. The S&P500 Index's constituents are both highly liquid and heavily traded, with the Index comprising of 504 common stocks due to certain companies listing two share classes (S&P Dow Jones Indices, 2022). With a minimum market capitalisation requirement of \$14.6B, the S&P500 Index represents approximately 80% of the American market, with Index constituents subject to quarterly review and rebalancing if necessary (S&P Dow Jones Indices, 2022).

For the 84-month period between January 2015 and December 2021, monthly data for the Total Return Index of the constituent securities was obtained from the Bloomberg Terminal. The Bloomberg Terminal was also used to collect monthly data related to the short interest ratio and short float ratio, earnings yield, momentum variables, such as the relative share price momentum and relative strength index, market capitalisation, beta, and the risk-free rate, proxied by the one-month United States Treasury Bill. Although some explanatory variables were calculated by Bloomberg, the accuracy of these inputs can be reasonably assumed given the financial database's reputation. Certain model inputs will be further discussed in the methodology section, together with their calculations.

Microsoft Excel was used to clean the dataset of missing values, restructure it from a wide to a long data format, and lag the explanatory variables. The study was conducted on an unbalanced dataset, with the number of firms present in the sample ranging from 415 to 500 over the period, resulting in 38087 observations. The dataset was then imported into RStudio, where all calculations, regression analyses, graphs, and tables were produced.

4 Methodology

4.1 Return Calculations

While Schindler, (2015) manually adjusted closing prices, this study utilised the Total Return Index (TRI) for constituent securities. The TRI provides a more accurate representation of a firm's equity value than the closing market price, as it is amended for corporate actions such as dividends, stock splits, rights offerings, and mergers (Ackert and Athanassakos, 2005). Prior to conducting the regression analyses, the monthly raw returns of constituent securities were calculated according to the following equation:

$$Return\ (R_{i,t}) = \left[\frac{TRI_{i,t}}{TRI_{i,t-1}} - 1\right] * 100 \tag{1}$$

Where:

$R_{i,t}$ = Monthly percentage return of firm i at time t

$TRI_{i,t}$ = Total return index of firm i at time t

$TRI_{i,t-1}$ = Total return index of firm i at time $t-1$

Note: Percentages are scaled by 100 to allow for easier interpretation of regression coefficients.

Diamond and Verrecchia, (1987) and Desai et al., (2002) argued that the performance of a stock is better measured by excess return as opposed to the raw return calculated in Equation 1, as prevailing market conditions such as the risk-free rate are incorporated. Thus, excess returns are computed as per the following equation:

$$Excess\ Return\ (ER_{i,t}) = Return_{i,t} - [Rf_t * 100] \tag{2}$$

Where:

$ER_{i,t}$ = Monthly percentage excess return of firm i at time t

$R_{i,t}$ = Monthly return of firm i at time t. *Equation 1*

Rf_t = Risk-free rate at time t. *Proxied by the one-month United States Treasury Bill Rate at time t*

Note: Percentages are scaled by 100 to allow for easier interpretation of regression coefficients.

Therefore, Equation 2 serves as the dependent variable in the regression analyses to investigate the cross-sectional relationship between short interest and excess returns of the constituent securities of the S&P500 Index.

4.2 Short Interest Ratio

The level of short interest was previously only reported by member firms to the United States Security Exchanges on the 15$^{\text{th}}$ calendar day of each month (Ackert and Athanassakos, 2005). As of September 2007, short interest reporting requirements were altered by the Financial Industry Regulatory Authority (FINRA) to semimonthly, with short positions disclosed at both the middle and end of each month, per Rule 4560 (Davis, 2012). For consistency with modern research, this study utilised monthly short interest data (Nezafat et al., 2019; Zhu et al., 2019).

The short interest ratio (SIR), serving as a proxy for short interest, is a quantitative indicator which represents the number of shares shorted for a particular security in relation to the average daily trading volume (ADTV) (Brent et al., 1990). Diether et al., (2009) argued that the SIR is frequently interchanged with the days to cover ratio, which denotes the average number of days required for an investor to cover their short position by repurchasing borrowed stock on the open market. Academics have utilised two differing measures of short interest in prior research, with Desai et al., (2002) and Asquith et al., (2005) suggesting that the ratio be calculated as the total number of shares sold short divided by the total number of shares outstanding. In contrast, Ackert and Athanassakos, (2005) and Schindler, (2015) defined the SIR as the total number of shares sold short divided by $ADTV$, as depicted by the following equation:

$$Short\ Interest\ Ratio\ (SIR_{i,t}) = \frac{SI_{i,t}}{ADTV_{i,t}} \tag{3}$$

Where:

$SIR_{i,t}$ = Short interest ratio of firm i at time t

$SI_{i,t}$ = Short interest of firm i at time t

$ADTV_{i,t}$ = Average daily trading volume of firm i at time t

Ackert and Athanassakos, (2005) argued that $ADTV$ is a preferred divisor, as it is representative of actual trading activity based on investors' decisions to buy and sell securities, which makes it more informative than short interest defined in relation to shares outstanding. Due to short-sellers frequently holding positions on a short-term basis, with an average days to cover ratio ranging between three and five days, Boehmer and Wu, (2013) suggested that $ADTV$ is favoured over shares outstanding as it accounts for liquidity. Schindler, (2015) further stated that $ADTV$ results in a standardisation between securities with differing trading volumes, allowing for short positions across securities to be better compared. Therefore, Equation 3 will be used as a proxy for short interest in this study.

Prior empirical research has also focused on the changes in SIR, as it provides insight into the revisions of speculative investors' expectations (Seneca, 1967; Hanna, 1976; Hurtado-Sanchez, 1978; Brent et al., 1990; Zhu et al., 2019). Instead of utilising the simple difference between the SIR in a given month relative to the preceding month ($\Delta SIR = SIR_{i,t} - SIR_{i,t-1}$), and in accordance with Zhu et al., (2019), monthly changes in the SIR are measured as per the following equation:

$$Percentage\ Change\ in\ SIR\ (\%\Delta\, SIR_{i,t}) = \left[\frac{SIR_{i,t}}{SIR_{i,t-1}} - 1\right] * 100 \tag{4}$$

Where:

$\%\Delta\, SIR_{i,t}$ = Monthly percentage change of the short interest ratio of firm i at time t

$SIR_{i,t}$ = Short interest ratio of firm i at time t

$SIR_{i,t-1}$ = Short interest ratio of firm i at time $t-1$

Note: Percentages are scaled by 100 to allow for easier interpretation of regression coefficients.

Equation 4 is a preferred measure as it captures more information than the simple difference approach. For instance, if Stock X's SIR increased from 1% to 3% and Stock Y's SIR increased from 2% to 4% over a one-month period, the simple difference would indicate that both securities experienced a 2% change. However, based on Equation 4, Stock X experienced an increase in the SIR of 200%, whereas Stock Y experienced a 100% increase in the SIR. Therefore, Stock X was subjected to more severe short selling than Stock Y, which the simple difference approach would not account for. This increased short-selling activity could signal changes in firm-specific fundamentals amid the dynamic economic environment, which may influence the future share price (Zhu et al., 2019). Thus, Equation 4 will be used to investigate the effect of changes in the short interest ratio on subsequent stock returns. Zhu et al., (2019) further stated that although changes in the SIR contain incremental predictive information regarding return predictability in the cross-section, the effect is not subsumed by current short interest levels and other widely known determinants of share return.

4.3 Short Float Ratio

Various studies have defined short interest relative to the total number of shares outstanding (Desai et al., 2002; Asquith et al., 2005; Schindler, 2015). While some authors have applied this definition as their primary proxy for short interest in prior research, others have utilised it as a robustness test for the traditional measure, as displayed in Equation 3. Schindler, (2015) argued that while short interest defined in relation to $ADTV$ is preferred, it does not account for illiquid securities which may be short sold in large volumes by an institutional investor who potentially possesses insider information. As a result, the following equation can be used as an alternative proxy for short interest:

$$Short\ Float\ Ratio\ (SFR_{i,t}) = \left[\frac{SI_{i,t}}{Equity\ float_{i,t}}\right] * 100 \tag{5}$$

Where:

$SFR_{i,t}$ = Percentage of firm i shorted at time t

$SI_{i,t}$ = Short interest of firm i at time t

$Equity\,Float_{i,t}$ = Total number of shares available to the public of firm i at time t

Shares available to the public is calculated as the total number of shares outstanding subtracted by shares held by insiders and shares which are stagnant.

Note: Percentages are scaled by 100 to allow for easier interpretation of regression coefficients.

Since the numerator is identical for each specification, monthly data is also used for this alternative definition of short interest (Nezafat et al., 2019). While the traditional measure is favoured for the reasons discussed by Ackert and Athanassakos, (2005), Boehmer and Wu, (2013), and Schindler, (2015) in Section 4.2, according to Asquith et al., (2005) the preferred divisor is partially dependent on the aim of the study. If short interest is thought to be indicative of future buying pressure when short positions are covered, as suggested by Seneca, (1967) and Mayor, (1968), the SIR in Equation 3 is arguably a better proxy (Asquith et al., 2005). In contrast, if short interest is seen as a reflection of firm-specific information held by informed investors, as stated by Boehmer et al., (2008), Akbas et al., (2013), and Callen and Fang, (2015), the SFR in Equation 5 is seen as a preferred measure (Asquith et al., 2005). For completeness, Equation 5 will be used as an alternative proxy for short interest in this paper.

Although the divisor between the two ratios differ, Asquith et al., (2005) argue that the two measures have a strong positive correlation. Schindler, (2015) further noted that the manner in which short interest is defined is often irrelevant, as the relationship between both measures and subsequent return remains largely unchanged. For consistency and the reasons cited in Section 4.2, the monthly percentage change in SFR will also be examined, as per the following equation:

$$Percentage\ Change\ in\ SFR\ (\%\Delta\, SFR_{i,t}) = \left[\frac{SFR_{i,t}}{SFR_{i,t-1}} - 1\right] * 100 \tag{6}$$

Where:

$\%\Delta\, SFR_{i,t}$ = Monthly percentage change of the short float ratio of firm i at time t

$SFR_{i,t}$ = Short float ratio of firm i at time t

$SFR_{i,t-1}$ = Short float ratio of firm i at time $t-1$

Note: Percentages are scaled by 100 to allow for easier interpretation of regression coefficients.

Changes in SFR may provide insight into changes in firm-specific fundamentals, which may impact future excess return, according to the reasoning of Asquith et al., (2005) and Zhu et al., (2019). Hence, Equation 6 will form part of the explanatory variables utilised in this paper.

4.4 Controlling for Other Variables

Following prior literature, this study controls for other variables which are known to have predictive ability with regard to the cross-section of future share returns. While Asquith et al., (2005) and Zhu et al., (2019) examined the book-to-market ratio, this study, in accordance with Fama and French, (1993), and Dechow et al., (2001) utilised the earnings yield ($EY_{i,t}$) as a measure of value, as it is both consistent in ranking and mitigates a low base effect. Dechow et al., (2001) noted that firms that displayed low fundamental ratios, particularly in relation to earnings, were frequently shorted as lower future returns were expected, however, in instances where these ratios mean-reverted, short-sellers were found to cover their positions.

To control for the potential of a price reversal ($PXREV_{i,t}$) effect, Angel et al., (2003) suggested including a short-term momentum variable, hence, the following equation:

$$Relative\ Strength\ Index\ 30\ Day\ (PXREVi,t) = 100 - \left[\frac{100}{1 + \frac{Average\ Up_{i,t}}{Average\ Down_{i,t}}} \right] \tag{7}$$

Where:

$PXREV_{i,t}$ = Non-trending indicator measuring the *30* day momentum of stock i at time t

$Average\ Up_{i,t}$ = Average of all day-on-day changes where stock i closed up at time t

$Average\ Down_{i,t}$ = Average of all day-on-day changes where stock i closed down at time t

The ratio in Equation 7 was calculated by Bloomberg and serves as a proxy of the price reversal effect in this study. The figure is for a one-month period, where a reading between *20* and *30* indicates a security is potentially oversold and is likely to experience an upward correction. Similarly, a reading between *70* and *80* indicates a security is potentially overbought and likely to experience a downward correction.

Barber et al., (1999) argued that in instances where a firm has previously experienced extreme return performance, controlling for this performance is necessary. Desai et al., (2002), Asquith et al., (2005), and Zhu et al., (2019) included a momentum factor that controlled for the prior six-month return, as stocks that were heavily shorted were found to have previously experienced a large upward run in price. The Relative Share Price Momentum ($MOM_{i,t}$), will serve as a proxy for momentum in this paper. The ratio is calculated by Bloomberg as the percentage change over the prior six months in the one-month moving average of the stock price relative to a benchmark index, which, for this particular study, is the S&P500 Index.

Smith, (1968) found evidence of a size effect, as short interest had a substantially larger impact on future returns for firms that had a lower supply of shortable shares. Both Asquith et al., (2005) and Boehme et al., (2006) noted that smaller market capitalisation stocks with high levels of short interest consistently underperformed relative to the market. Following prior research, the natural logarithm of market capitalisation ($Size_{i,t}$) serves as a proxy for size in this paper (Smith, 1968; Ackert and Athanassakos, 2005; Asquith et al., 2005; Zhu et al., 2019). Market capitalisation is calculated by Bloomberg by multiplying a firm's outstanding shares by the current market price, which denotes the total market value of a company in US dollars.

Baron and McDonald, (1973), Brent et al., (1990), and Werner, (2010) reasoned that short selling is also frequently motivated by arbitrage and hedging. Baron and McDonald, (1973) suggested that the explanatory variable, beta, is often used to determine the existence of arbitrage and hedging in the market. The authors further noted that there is a positive correlation between short interest levels and beta (Baron and McDonald, 1973). Brent et al., (1990) argued that high beta shares are desirable vehicles for arbitrage and hedging due to their high correlation with the market. Thus, beta ($Beta_{i,t}$) will act as a proxy for arbitrage and hedging in this study.

4.5 Derived Hypothesis

The majority of modern academic research advocates for the speculative view, where short-sellers intend to profit from declining stock prices (Callen and Fang, 2015; Nezafat et al., 2019; Zhu et al., 2019). However, this study, in line with Mayor, (1968), Pinches, (1970), and Ackert and Athanassakos, (2005) instead considers the symmetry of outcomes. Thus, the null hypothesis:

H_0 : There is no relationship between short interest and future excess stock return.

Failure to reject H_0 thus provides evidence supporting the efficient market hypothesis (Pinches, 1970). If H_0 is rejected due to a statistically significant negative coefficient on short interest, this provides evidence supporting the speculative view (Callen and Fang, 2015). In contrast, if H_0 is rejected due to a statistically significant positive coefficient, this provides support for the contrarian view (Schindler, 2015).

5 Descriptive Statistics

5.1 The Trend of Short Interest

Figure 1 displays the monthly trend of the equal-weighted average of short interest, defined in relation to both *ADTV* and *Equity Float* for the period between January 2015 and December 2021. Although Rapach et al., (2016) utilised the value-weighted average of short interest, where higher *(lower)* market capitalisation stocks received a larger *(smaller)* weighting, both Desai et al., (2002) and Asquith et al., (2005) argued that the effect of short selling in large market capitalisation stocks is often less significant.

Figure 1:

Short Interest Over Time

As a result, the equal-weighted average of short interest depicted above, better emphasises segments of the shorting market that are likely to be more active (Desai et al., 2002). The distribution of the short interest measures exhibits four distinct characteristics. Firstly, since each short interest measure is based on the same numerator (SI), the difference in distributions is solely attributable to the divisors. The SIR is substantially more volatile than the SFR due to the volatility in $ADTV$, which motivated Desai et al., (2002) and Asquith et al., (2005) to analyse the trend of short interest by exclusively examining the SFR, defined in relation to $Equity\ Float$, as it is less distorted by trading volume.

Secondly, at the height of short selling in 2016, the monthly average of the SIR was only 5.2 days, whereas the SFR was only 4.15%, indicating that firms within the sample generally have low levels of short interest (Asquith et al., 2005). Thirdly, concerning SFR, there is no distinct seasonality effect, as short positions do not continuously peak at year-end, indicating that shorting is not used for taxation deferment purposes (Brent et al., 1990).

Lastly, in contrast to Desai et al., (2002) and Asquith et al., (2005) who found short interest levels to be consistently increasing, this study reports an increasing trend until 2016, when short interest levels, defined by SFR, then steadily declined until the end of the sample period. The decline is possibly due to two reasons, with Demirer et al., (2019) first proposing that the S&P500 is currently categorised as a bubble environment due to the disconnect between market prices and fundamentals, which has disincentivised short selling. Hence, short-sellers often require a catalytic event to initiate the convergence between price and fundamental value, before placing their bets (Demirer et al., 2019).

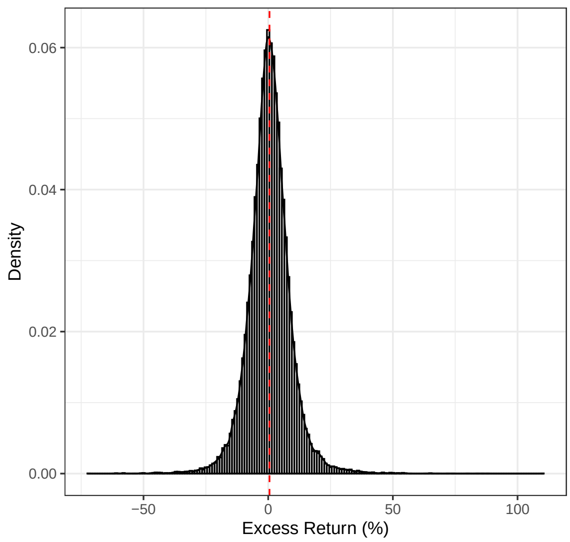
Interestingly, amid the COVID-19 pandemic, short selling, which was not banned in the United States, briefly increased in 2020 before steeply declining, indicating that short-sellers do not exert further downward pressure in extreme market conditions (Fotak et al., 2009). The pandemic, together with the counter-hegemonic financial efforts of retail investors, as described by Chohan, (2021) in relation to the GameStop short squeeze of 2021, was cited as a second possible explanation for the decline. As a result, professional investors have become increasingly cautious about shorting firms with extreme levels of short interest as retail investors could potentially squeeze shares regardless of fundamentals, which has caused a risk and reward imbalance with regards to short positions, leading investors to instead hedge their positions with futures (Anand and Pathak, 2022).

5.2 Variable Histograms

Following Schindler, (2015) the variables within the dataset were then analysed to ensure regression suitability. Through the utilisation of histograms the distributions of the variables were checked for normality. The dependent variable, excess return, is displayed in Figure 2 below. The variable is approximately normally distributed, which Diamond and Verrecchia, (1987) argued is a result of short selling being permitted throughout the entire sample period, as it allows negative information to be incorporated into the market and prevents bullish investors from skewing the distribution to the left. Siciliano and Ventoruzzo, (2020) found that had short selling restrictions been imposed during the pandemic, as they were for selected stocks in Austria, Belgium, France, Greece, Italy, and Spain, the excess returns would have been lower due to lower liquidity and information asymmetry, resulting in a positive skew and causing the outcome the bans aimed to prevent.

Figure 2:

Histogram of Excess Return



Both measures of short interest were then analysed to determine their distributions, with the histograms of SIR and SFR displayed in Figure 3 and Figure 4 respectively. Both figures indicate that each short interest measure follows a right-skewed distribution.

Figure 3:

Histogram of SIR

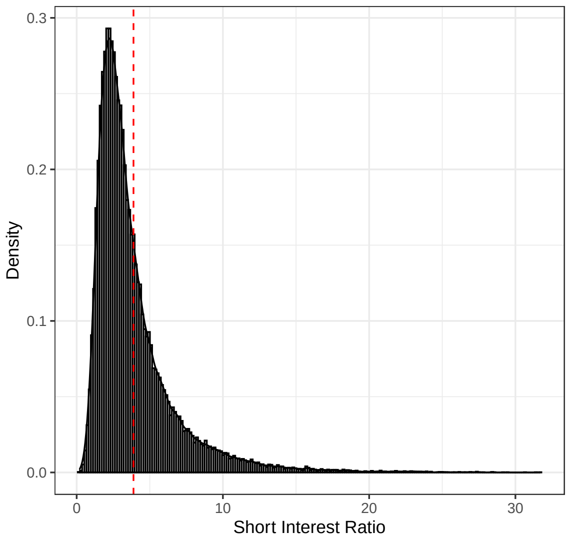
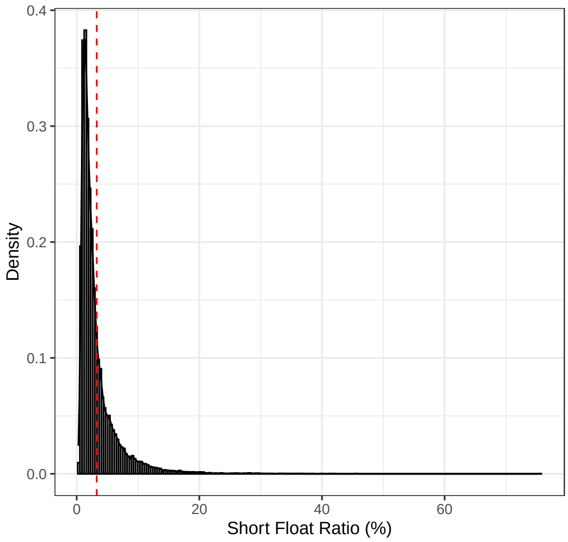


Figure 4:

Histogram of SFR



These positively skewed distributions are due to the majority of firms in the sample having an insignificant number of shares shorted, whereas only a few firms are subjected to substantial short selling (Asquith et al., 2005; Nezafat et al., 2019). Consistent with Ackert and Athanassakos, (2005), Asquith et al., (2005), and Schindler, (2015) the natural logarithm was used to transform both the SIR and SFR variables to LN_SIR and LN_SFR , the distributions of which are displayed in Figure 5 and Figure 6 respectively.

Figure 5:

Histogram of Natural Log of SIR

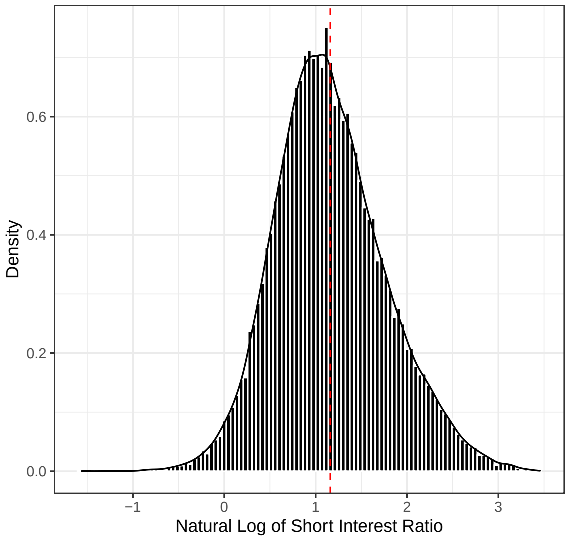
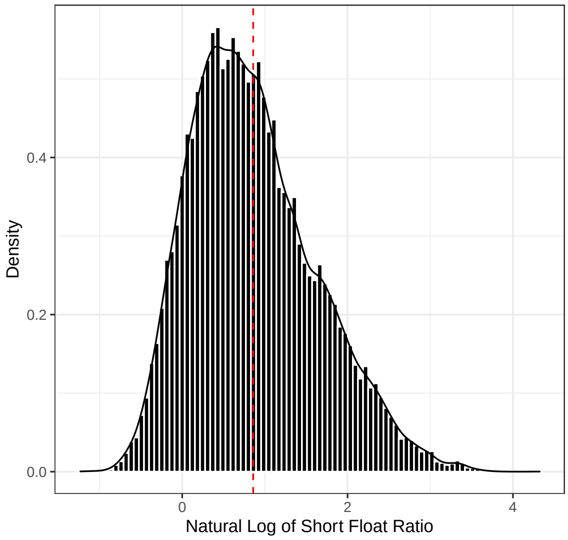


Figure 6:

Histogram of Natural Log of SFR



As expected, the transformed short interest variables are more normally distributed and will form part of the explanatory variables in the regression analyses (Schindler, 2015). The histograms for the remainder of the variables are displayed in Appendix A. Similar to the short interest measures, the market capitalisation variable was also skewed to the right. Hence, in accordance with prior research the natural logarithm was used to transform this variable as well (Asquith et al., 2005; Schindler, 2015; Zhu et al., 2019).

5.3 Summary Statistics

Table 1 displays the summary statistics for the variables discussed in Section 4 and Section 5.2. There are 38087 observations for each variable over the 2015 to 2021 sample period. ER has a mean of 0.524%. A mean of excess return similar to zero was also reported by both Albert et al., (1997) and Schindler, (2015) who investigated the NYSE and NASDAQ, respectively. The mean SIR of 3.885 days is consistent with the findings of Boehmer and Wu, (2013), who reported an average days-to-cover ranging between three and five days on the aforementioned exchanges. As graphically depicted in Figure 3, the SIR is highly skewed, with a skewness of 2.659, which was the motive for logging the variable. As shown by LN_SIR below, the natural logarithm substantially reduces the variables' skewness.

Table 1:

Summary Statistics

	Mean	Standard Deviation	Skewness	Kurtosis
ER (%)	0.524	8.609	0.306	9.184
SIR	3.885	2.871	2.659	13.769
LN_SIR	1.161	0.606	0.329	3.245
CHANGE_SIR (%)	7.985	47.344	2.807	32.425
SFR (%)	3.271	3.417	3.683	26.215
LN_SFR	0.858	0.759	0.574	3.010
CHANGE_SFR (%)	1.893	25.484	6.812	137.767
EY (%)	4.622	5.482	−8.044	224.607
PXREV	58.743	9.753	−0.686	6.835
MOM	1.402	18.139	1.936	20.570
MARKET_CAP	54077.240	123639.600	8.953	118.131
LN_MARKET_CAP	10.155	1.077	0.614	4.182
BETA	1.046	0.599	0.364	11.402

Number of observations = 38087

The $CHANGE\_SIR$ mean of 7.895% indicates that the SIR changes significantly month-to-month, however, this is most likely distorted by the volatility in trading volume (Zhu et al., 2019). The SFR mean of 3.271% indicates that the typical S&P500 Index firm has few outstanding shares shorted (Asquith et al., 2005). The skewness of the SFR measure further substantiates why it was logged, with the LN_SFR appearing more normally distributed. The $CHANGE\_SFR$ mean of 1.893% is substantially lower than the mean of the $CHANGE\_SIR$ and better depicts how short interest changes on a monthly basis (Zhu et al., 2019).

The majority of the variables in the dataset are moderately to highly skewed, and since kurtosis is a function of the prior, each variable is best described as a leptokurtic (Field, 2009). Leptokurtic distributions are not uncommon in financial data, as trading information generally appears in clusters as opposed to a continuous linear flow (Peters, 1996). The heavy-tails of the variables imply an increased probability of extreme outliers in the dataset, however, these outliers were not removed, as Au et al., (2009) suggested that at the extremities is where the effect of short selling is likely to be most prevalent. Nevertheless, the dataset utilised in this study is deemed reliable, as the summary statistics do not display any violations of the Ordinary Least Squares assumptions.

5.4 Correlation Matrix

In accordance with Schindler, (2015) and Rapach et al., (2016) the dataset was then checked for multicollinearity using Pearson's correlation matrix, as shown in Table 2 below. To ensure that cross-sectional correlations were reported, the correlation between each variable was calculated monthly and averaged over the sample period. As per Wooldridge, (2015), instances in which the correlation coefficient is greater than the absolute value of 0.2 indicate potential multicollinearity.

As expected, the short interest measures, together with their logarithmic transformations and monthly changes, are above the correlation threshold; hence, each measure will be utilised as an explanatory variable in separate regression analyses to avoid potential multicollinearity. From the correlation matrix, multicollinearity was identified between LN_MARKET_CAP, acting as a proxy for size, and the short interest measures. This is a result of short interest being scaled by $ADTV$ and $Equity\ Float$. Hence, in contrast to Ackert and Athanassakos, (2005), Asquith et al., (2005), and Zhu et al., (2019) the explanatory variable $Size_{i,t}$ was excluded from this study.

Table 2:

Correlation Matrix

	ER	SIR	LN_SIR	CHANGE_SIR	SFR	LN_SFR	CHANGE_SFR	EY	PXREV	MOM	MARKET_CAP	LN_MARKET_CAP	BETA
ER	1												
SIR	0.002	1											
LN_SIR	-0.005	**0.913**	1										
CHANGE_SIR	-0.007	**0.226**	**0.273**	1									
SFR	0.025	**0.653**	**0.591**	0.060	1								
LN_SFR	0.024	**0.685**	**0.705**	0.087	**0.872**	1							
CHANGE_SFR	-0.012	0.053	0.081	**0.407**	**0.215**	**0.222**	1						
EY	-0.022	-0.067	-0.069	-0.009	-0.138	-0.103	0.023	1					
PXREV	0.004	-0.054	-0.052	-0.000	-0.197	-0.019	-0.030	-0.171	1				
MOM	0.010	-0.014	-0.015	-0.020	-0.043	-0.032	-0.053	-0.123	0.198	1			
MARKET_CAP	-0.007	-0.193	**-0.248**	-0.191	**-0.215**	**-0.355**	-0.197	0.000	0.136	0.048	1		
LN_MARKET_CAP	-0.048	**-0.308**	**-0.343**	**-0.293**	**-0.384**	**-0.534**	**-0.227**	0.100	0.177	0.029	**0.739**	1	
BETA	0.038	-0.106	-0.123	0.004	0.053	0.067	0.001	0.057	-0.186	0.012	-0.021	-0.089	1

Correlations coefficients $> |0.2|$ *are in* ***bold*** *and represent potential multicollinearity (Wooldridge, 2015).*

The signs of the coefficients between ER and the short interest measures corroborate alternative views. Following the contrarian view and in accordance with Schindler, (2015), ER is positively correlated with both measures of short interest, SIR and SFR, and the transformed LN_SFR variable. In contrast, the coefficients of LN_SIR and changes in the short interest variables, $CHANGE\_SIR$ and $CHANGE\_SFR$, imply a negative correlation is present, following the speculative view, as found by Au et al., (2009), Akbas et al., (2013), and Callen and Fang, (2015). The relationship between excess return and the short interest measures, together with other well-known predictors of stock return, will be further investigated through regression analyses.

6 Fama-Macbeth (1973) Regression Approach

Aitken et al., (1998) and Desai et al., (2002) advocated for the calendar time portfolio approach, formed by Jaffe, (1974) and Mandelker, (1974), due to the methods' ability to detect non-monotonic relationships between variables. However, Akbas et al., (2013) argued against the approach, as a possible shortcoming was spurious relations occurring as a result of cross-sectional correlation in the residuals due to excluded variables. To address the potential bias within the standard error of estimate, and following Ackert and Athanassakos, (2005), Akbas et al., (2013), Callen and Fang, (2015), and Zhu et al., (2019) this study employs the Fama and MacBeth (1973) cross-sectional regression approach to analyse whether an economically and statistically significant relationship exists between the level of the short interest and future share returns over the sample period.

The regressions are estimated for each monthly period, where the monthly coefficient estimates are then averaged across time, with the significance levels determined by pooled *t-statistics*, as per the following equation:

$$t_k = \frac{\bar{b}_{k,t}}{\frac{\sigma_k}{\sqrt{T}}} \tag{8}$$

Where:

t_k = Pooled *t-statistic* of variable k

$\bar{b}_{k,t}$ = Average monthly coefficient estimate of an independent variable k

σ_k = Standard deviation of the coefficient estimate k

T = Number of monthly sample periods

6.1 Short Interest Ratio - Univariate Regression

In line with Schindler, (2015) and Zhu et al., (2019) a univariate analysis is first conducted, where the dependent variable, excess return, is regressed against the explanatory variable, the short interest ratio, which is lagged by one month, as per the regression below:

$$ER_{i,t} = \alpha + \beta_1 * SIR_{i,t-1} + \epsilon_t \tag{9}$$

Where:

$ER_{i,t}$ = Monthly excess return of stock i at time t. *Equation 2*

α = Regression constant

$SIR_{i,t-1}$ = Short interest ratio of stock at i at time $t-1$. *Equation 3*

The $SIR_{i,t-1}$ measure will also take on the following values:

$LN\_SIR_{i,t-1}$ = Natural logarithm of $SIR_{i,t-1}$

$\%\Delta\, SIR_{i,t}$ = Monthly percentage change in SIR of firm i at time t. *Equation 4*

ϵ_t = Regression error

6.2 Short Float Ratio - Univariate Regression

Following the recommendations of Ackert and Athanassakos, (2005) and Schindler, (2015), an alternative measure of short interest, defined by the short float ratio, is also examined:

$$ER_{i,t} = \alpha + \beta_1 * SFR_{i,t-1} + \epsilon_t \tag{10}$$

Where:

$SFR_{i,t-1}$ = Short float ratio of stock at i at time $t-1$. *Equation 5*

The $SFR_{i,t-1}$ measure will also take on the following values:

$LN\_SFR_{i,t-1}$ = Natural logarithm of $SFR_{i,t-1}$

$\%\Delta\, SFR_{i,t}$ = Monthly percentage change in SFR of firm i at time t. *Equation 6*

All other variables are interpreted in accordance with Regression 9

In contrast to Mayor, (1968), who stated that a non-zero coefficient is indicative of statistical significance, both Woolridge and Dickson, (1994) and Schindler, (2015) argued that the coefficient β_1 in both Regression 9 and Regression 10 must be significantly different from zero in order to be statistically significant since factors such as information do not remain constant.

6.3 Short Interest Ratio - Multiple Regression

The Fama and MacBeth (1973) method further allows for the significance of short interest measures to be simultaneously examined after controlling for other well-known predictors of share return. Figlewski and Webb, (1993) and Ackert and Athanassakos, (2005) argued for the variable *"option"* to be included, to indicate whether a company has stock options available, as both authors assumed that options would mitigate the effect that short interest levels have on excess returns. This paper examines the S&P500 Index, which is made up of the largest 500 companies in the United States as weighted by market capitalisation, and all constituent firms are optioned, allowing the variable to be excluded. Instead, the lagged explanatory variables, excluding $Size_{i,t}$, described in Section 4.4, are introduced, as depicted by the multiple regression below:

$$ER_{i,t} = \alpha + \beta_1 * SIR_{i,t-1} + \beta_2 * EY_{i,t-1} + \beta_3 * PXREV_{i,t-1} + \beta_4 * MOM_{i,t-1} + \beta_6 * Beta_{i,t-1} + \epsilon_t \tag{11}$$

Where:

$ER_{i,t}$ = Monthly excess return of stock i at time t. *Equation 2*

α = Regression constant

$SIR_{i,t-1}$ = Short interest ratio of stock at i at time $t-1$. *Equation 3*

The $SIR_{i,t-1}$ measure will also take on the following values:

$LN\_SIR_{i,t-1}$ = Natural logarithm of $SIR_{i,t-1}$

$\%\Delta\, SIR_{i,t}$ = Monthly percentage change in SIR of firm i at time t. *Equation 4*

$EY_{i,t-1}$ = Earnings yield of stock i at time $t-1$

$PXREV_{i,t-1}$ = *30* day momentum of stock i at time $t-1$. *Equation 7*

$MOM_{i,t-1}$ = *6* month momentum of stock i at time $t-1$

$Beta_{i,t-1}$ = Beta of stock i at time $t-1$

ϵ_t = Regression error

6.4 Short Float Ratio - Multiple Regression

For consistency and to ensure the robustness of results, the alternative measure of short interest, utilised by Ackert and Athanassakos, (2005) and Schindler, (2015), is examined in the multiple regression framework:

$$ER_{i,t} = \alpha + \beta_1 * SFR_{i,t-1} + \beta_2 * EY_{i,t-1} + \beta_3 * PXREV_{i,t-1} + \beta_4 * MOM_{i,t-1} + \beta_6 * Beta_{i,t-1} + \epsilon_t \tag{12}$$

Where:

$SFR_{i,t-1}$ = Short float ratio of stock at i at time $t-1$. *Equation 3*

The $SFR_{i,t}$ measure will also take on the following values:

$LN\_SFR_{i,t-1}$ = Natural logarithm of $SFR_{i,t-1}$

$\%\Delta\, SFR_{i,t}$ = Monthly percentage change in SFR of firm i at time t. *Equation 4*

All other variables interpreted in accordance with Regression 11

Due to the focus of this study centered on determining the explanatory power of short interest on future excess returns, by first conducting univariate regression analyses *(Regression 9 and Regression 10)* and subsequently controlling for firm-specific fundamentals in the multiple regression analyses *(Regression 11 and Regression 12)*, the coefficient of the short interest measures in both specifications can be compared. According to Akbas et al., (2013), if the regression coefficient of short interest reduces in magnitude once these explanatory variables are introduced, short-sellers ability to predict future share returns is partially explained by the information that these firm fundamentals portray.

7 Empirical Results

7.1 Univariate Analyses of the Short Interest Ratio

The baseline regression models of the univariate analyses of the SIR *(Regression 9)* presented in Table 3 indicate that no relationship is present between the SIR measures and future excess return in the cross-section. The coefficient of the SIR in Model 1 is positive, in line with Schindler, (2015), however, it is statistically insignificant.

In contrast, the coefficients of both the LN_SIR presented in Model 2 and the $CHANGE\_SIR$ in Model 3 are negative, in accordance with Ackert and Athanassakos, (2005) and Zhu et al., (2019) respectively. However, unlike the authors, neither variable is statistically significant. Although the R^2 of each model is large, the insignificance indicates that the short interest ratio is *nonpredictive* with regard to future excess return. Thus, we fail to reject the null hypothesis (H_0). The results of Table 3 hence provide support for the efficient market hypothesis, as found by Mayor, (1968) and Smith, (1968) in their empirical analyses of short interest. According to

Pinches, (1970), the main characteristic of an efficient market implies that, at any time, the market price of an asset is expected to closely represent the asset's intrinsic value. Therefore, the positive and negative implications of short interest should already be inherent in a security's trading price; hence, the level of short interest would have no impact on future excess return (Caster and Vu, 1987).

Table 3:

Fama-Macbeth Univariate Analyses of SIR

	Dependent variable:		
	ER		
	(1)	(2)	(3)
SIR	0.011 (0.448)		
LN_SIR		−0.036 (-0.315)	
CHANGE_SIR			−0.002 (-1.220)
Constant	0.457 (0.809)	0.548 (0.924)	0.459 (0.858)
Observations	38,087	38,087	38,087
R^2	0.316	0.317	0.315

Significance Levels: *p<0.1; **p<0.05; ***p<0.01

Numbers in brackets represent two-tailed t-statistics

7.2 Multiple Regression Analyses of the Short Interest Ratio

Following prior research, four explanatory variables were then introduced to the baseline models in Table 3 *(Regression 11)*, the results of which are displayed in Table 4 below. In contrast to Akbas et al., (2013) who found that the coefficient of the short interest ratio declined once explanatory variables were introduced, the coefficient of the SIR in Model 4 instead increased in magnitude relative to Model 1, indicating that a short-sellers ability to predict returns is not explained by predictions of firm-specific fundamentals.

Although the significance of the coefficient of the SIR increased relative to Model 1, it remained statistically insignificant. Similarly, the coefficients of both the LN_SIR in Model 5 and the

$CHANGE\_SIR$ in Model 6 remained statistically insignificant. However, the coefficient of the LN_SIR experienced a sign change, while the coefficient of the $CHANGE\_SIR$ marginally decreased in magnitude, relative to Model 2 and Model 3, respectively.

Using a two-tailed test, the negative coefficient of EY in Model 4 is statistically significant at the 10% level, whereas the coefficients of the variable in Models 5 and 6 are statistically significant at the 5% level. Therefore, a 1% increase in EY in Model 4, (5), and [6] is associated with a decrease in excess returns of 4.4 *basis points* (*bp*), (4.6*bp*), and [4.7*bp*] in the subsequent month.

Table 4:

Fama-Macbeth Multiple Regression Analyses of SIR

	Dependent variable:		
	ER		
	(4)	(5)	(6)
SIR	0.024 (1.224)		
LN_SIR		0.016 (0.176)	
CHANGE_SIR			−0.001 (-0.966)
EY	−0.044* (-1.893)	−0.046** (-1.973)	−0.047** (-1.983)
PXREV	0.013 (0.788)	0.012 (0.760)	0.012 (0.736)
MOM	0.001 (0.269)	0.001 (0.288)	0.001 (0.225)
BETA	0.531 (1.576)	0.525 (1.567)	0.525 (1.558)
Constant	−0.744 (-0.705)	−0.623 (-0.579)	−0.617 (-0.596)
Observations	38,087	38,087	38,087
R^2	0.407	0.408	0.406

Significance Level: *p<0.1; **p<0.05; ***p<0.01

Numbers in brackets represent two-tailed t-statistics

While the negative coefficients of EY are contrary to the findings of Dechow et al., (2001), who reported a statistically significant positive coefficient, Fama and French, (1993) argued that the unexpected sign is likely a result of the substantial variability of earnings, which are directly related to future returns.

The positive coefficients of $PXREV$ and MOM for all models in Table 4 are consistent with the empirical evidence of Angel et al., (2003) and Akbas et al., (2013), respectively. However, unlike the aforementioned authors, the price reversal and momentum variables are statistically insignificant. While the coefficients for $Beta$ are positive for all models, as found by Baron and McDonald, (1973) and Brent et al., (1990), the coefficients for the variable are also insignificant. Although the R^2 increased for all models relative to those in Table 3, the statistical insignificance of the coefficients of the short interest measures in Table 4 indicated that even after controlling for fundamentals, the short interest ratio has no predictive power with regards to future excess return (Pinches, 1970). Thus, we fail to reject the null hypothesis (H_0) based on the results of Models 4, 5, and 6.

7.3 Univariate Analyses of the Short Float Ratio

Following Desai et al., (2002), Asquith et al., (2005), and Schindler, (2015) and to ensure the reliability of results, short interest was also defined in relation to $Equity\ Float$. For consistency, the univariate analyses of the SFR *(Regression 10)* are presented in Table 5 below. The coefficient of the SFR in Model 7 is larger than that of the SIR in Model 1, both in magnitude and significance. However, the t-statistic indicates that the coefficient remained narrowly insignificant. Hence, based on Model 7, we fail to reject the null hypothesis (H_0), indicating that the short float ratio has no predictive power with regards to future excess return.

Unlike Schindler, (2015) who argued that the empirical results were indifferent regardless of the manner in which short interest was defined, the coefficients of the LN_SFR and the $CHANGE\_SFR$ in Models 8 and 9 substantially differ from the LN_SIR and the $CHANGE\_SIR$ in Models 2 and 3, respectively. The coefficient of the LN_SFR in Model 8 is positive and statistically significant at the 10% level; hence, a 1% increase in the LN_SFR is associated with a 27.5*bp* increase in excess return in the subsequent month. Thus, we reject the null hypothesis (H_0) based on Model 8, as the LN_SFR has significant positive predictive power with regard to future excess return. As a result, the LN_SFR is therefore seen as a *bullish indicator* for contrarian investors, supporting the findings of Schindler, (2015). In contrast, the coefficient of the $CHANGE\_SFR$ is negative and statistically significant at the 5% level. Therefore, a 1% increase in the $CHANGE\_SFR$ is associated with a 0.4bp decrease in excess return in the subsequent month.

Table 5:

Fama-Macbeth Univariate Analyses of SFR

	Dependent variable:		
	ER		
	(7)	(8)	(9)
SFR	0.057 (1.639)		
LN_SFR		0.275* (1.904)	
CHANGE_SFR			−0.004** (-2.014)
Constant	0.300 (0.603)	0.268 (0.552)	0.491 (0.940)
Observations	38,087	38,087	38,087
R^2	0.328	0.326	0.315

Significance Level: *p<0.1; **p<0.05; ***p<0.01

Numbers in brackets represent two-tailed t-statistics

Thus, the null hypothesis (H_0) is also rejected based on Model 9, as the $CHANGE\_SFR$ has significant negative predictive power with regards to future excess return. The results of Model 9 support the findings of Zhu et al., (2019), who argued that changes in the short interest measure act as a *bearish indicator*, providing insight into the revisions of speculative investors' expectations.

7.4 Multiple Regression Analyses of the Short Float Ratio

The four explanatory variables are then introduced into the baseline regression models of the SFR in Table 5 *(Regression 12)*, the results of which are reported in Table 6 below. The coefficient of the SFR in Model 10 is significant at the 10% level, such that a 1% increase in the SFR is associated with a 4.3bp increase in excess return in the subsequent month. The SFR thus serves as an economically and statistically significant *bullish indicator* for investors in terms of future excess return, supporting the findings of Schindler, (2015). The statistical significance of the LN_SFR and the $CHANGE\_SFR$ have reversed relative to Models 8 and 9, such that the LN_SFR is significant at the 5% level, whereas the $CHANGE\_SFR$ is significant at the 10% level. Although the signs for each of the coefficients of the short float measures have remained unchanged, the magnitude of their coefficients have declined relative to the univariate analyses in Table 5.

Table 6:

Fama-Macbeth Multiple Regression Analyses of SFR

	Dependent variable:		
	ER		
	(10)	(11)	(12)
SFR	0.043* (1.657)		
LN_SFR		0.196** (2.027)	
CHANGE_SFR			−0.003* (-1.681)
EY	−0.041* (-1.748)	−0.044* (-1.864)	−0.046** (-1.961)
PXREV	0.016 (0.984)	0.016 (0.993)	0.011 (0.700)
MOM	0.002 (0.351)	0.001 (0.272)	0.001 (0.261)
BETA	0.494 (1.482)	0.486 (1.459)	0.508 (1.511)
Constant	−0.967 (-0.933)	−0.985 (-0.944)	−0.549 (-0.532)
Observations	38,087	38,087	38,087
R^2	0.412	0.410	0.406

Significance Level: *p<0.1; **p<0.05; ***p<0.01

Numbers in brackets represent two-tailed t-statistics

This corroborates the findings of Akbas et al., (2013), as short sellers' ability to predict future excess returns is partially explained by their ability to predict firm-specific fundamentals. A 1% increase in the LN_SFR is associated with a 19.6bp increase in excess return in the subsequent month, which is a decline from that in Model 8. Similarly, a 1% increase in $CHANGE\_SFR$ is associated with a 0.3bp decrease in excess return in the subsequent month.

Thus, we reject the null hypothesis (H_0) due to the high R^2 and the statistical significance of the coefficients of the short float measures in Models 10, 11, and 12, with the interpretation of the signal of the latter two models synonymous with those in Table 5. The signs of the coefficients of EY in Table 6 remained unchanged relative to those in Table 4. However, the statistical significance of EY in Model 11 is now significant at the 10% level as opposed to the 5% level in Model 5. Thus, a 1% increase in EY in Model 10, (11), and [12] is associated with a decrease in excess returns of 4.1bp, (4.4bp), and [4.6bp] in the subsequent month. The interpretation of $PXREV$, MOM and $Beta$ also follows those described in Table 5.

The positive coefficients of the SFR in Model 10 and the LN_SFR in Model 11 are contrary to the findings of the majority of recent empirical research, which has reported a statistically significant negative relationship between short interest and future excess return in the cross-section (Callen and Fang, 2015; Nezafat et al., 2019; Zhu et al., 2019). The differences in significance between the SIR measures and the SFR measures indicate that the short float ratio contains informational content surrounding firm-specific information, whereas the short interest ratio is not indicative of future buying pressure when short positions are covered (Asquith et al., 2005).

8 Tests for Robustness

While the inclusion of an alternative measure of short interest simultaneously served as a test for robustness, the Fama and MacBeth (1973) regression approach was also conducted on a balanced dataset, since there is no set preference between utilising balanced or unbalanced data in empirical analyses. Due to Boehmer and Wu, (2013) finding a days-to-cover ratio of three to five days, similar to that reported in Table 1 *(3.885)*, and following Zhu et al., (2019), a final robustness test was conducted in which the short interest measures were not lagged, as the authors argued that short interest data is available to numerous investors, particularly institutional, mid-month. As a result, lagging short interest measures may weaken any relationship between short interest and future excess return. For consistency, the four explanatory variables were not lagged either.

8.1 Balanced Dataset

The histograms and summary statistics of the balanced dataset were largely unchanged and were hence excluded from the paper. The correlation matrix along with the results of the balanced dataset are reported in Appendix B. While the magnitude of the correlation coefficients varied slightly from those reported in Table 2, the signs of the coefficients remained unchanged. LN_MARKET_CAP was again removed from the Fama-Macbeth regression analyses due to potential multicollinearity, however, all other variables were included as outlined in Regressions 9, 10, 11, and 12. Although the coefficients of both the SIR and the SFR measures differed slightly in magnitude, the variables of interest remained statistically significant and retained their respective signs for all models in which significance was previously reported for the unbalanced results in Section 7.

In contrast to the unbalanced dataset, the balanced results indicate that the coefficient of EY is now statistically insignificant for all models. Interestingly, the coefficients of $Beta$ for the balanced dataset remained positive, but were statistically significant for all multiple regression analyses at the 10% level, supporting the findings of Baron and McDonald, (1973) and Brent et al., (1990). This, together with the statistically significant positive coefficients of the SFR in the multiple regression analyses in Table 11, indicates that short selling is commonly motivated by arbitrage and hedging (Brent et al., 1990). Corroborating the findings of Werner, (2010), the use of shorting for arbitrage and hedging purposes by hedge funds utilising market-neutral strategies has reduced the negative predictive power of short interest on future returns reported by the majority of modern research.

8.2 Non-lagged Dataset

The results of the Fama and MacBeth (1973) non-lagged regression analyses are reported in Appendix C. The coefficients of the univariate analyses of the SIR measures, which were previously insignificant in Table 3, are positive and statistically significant at the 1% level in Table 12, indicating that SIR serves as a *bullish indicator* with regard to future excess return. Asquith et al., (2005) argued that the positive coefficients of the SIR measures are representative of a latent demand for the stock, which will result in future upward price pressure when short positions are covered. Similarly, the coefficients of the SIR measures in the multiple regression analyses reported in Table 13 are positive and statistically significant at the 1% level, following the same interpretation discussed above.

The coefficients of all explanatory variables in Table 13 are statistically significant, with the interpretation of EY and $Beta$ consistent with that of Section 7.4 and 8.1, respectively. The positive coefficients of $PXREV$ in the non-lagged results is statistically significant at the 1% level. Fosback, (1993) and Angel et al., (2003) argued that the positive coefficient of the price reversal effect is indicative of investors becoming excessively bearish, at which point the oversold security then experiences an upward price correction. Although the coefficient of MOM is negative and significant at the 10% level, the momentum variables' coefficient is of little economic significance.

The univariate analyses of the SFR measures in Table 14 indicate that while the significance of the SFR and the LN_SFR remained unchanged relative to the lagged sample, the coefficient of the $CHANGE\_SFR$ is statistically insignificant. The coefficients of SFR and LN_SFR in the multiple regression analyses in Table 15 are significant at the 1% level, with the $CHANGE\_SFR$ experiencing a sign change relative to Table 6. The coefficients of the explanatory variables in Table 15, which are all statistically significant, follow the same interpretation previously discussed.

The coefficients of the non-lagged dataset indicate that the relationship between the short interest measures and future excess return increased in both magnitude and statistical significance relative to the results of the lagged data reported in Section 7. Thus, supporting the findings of Zhu et al., (2019), who argued that by lagging short interest data, any relationship between short interest and excess returns will be weakened, as the majority of institutional investors have access to short interest data semi-monthly.

9 Cumulative Returns to Short Interest Measures

9.1 Cumulative Returns to the Short Interest Ratio

Following Au et al., (2009) and Zhu et al., (2019), the equal-weighted cumulative correlations were calculated between the short interest measures and one-month forward excess returns. As shown in Figure 7 below, a trading strategy based on the LN_SIR and the $CHANGE\_SIR$ was generally unprofitable over the sample period. Similarly, with regards to the SIR, superior investment timing was necessary, as it too was unprofitable between 2017 and mid-2018, as well as towards the end of 2021. The results of the univariate analyses of the SIR in Table 3 suggest that the market is efficient. Thus, Baron and McDonald, (1973) argue that speculative short-sellers are unable to consistently predict market movements and constituent securities price changes, which, when combined with the general upward trend of the S&P500 Index over the sample period, results in the returns to the SIR measures being negative in most instances.

Figure 7:

Cumulative Correlation - Short Interest Ratio and Excess Return

9.2 Cumulative Returns to the Short Float Ratio

The cumulative correlations between the short float ratio and one-month forward excess returns were then calculated, and are displayed in Figure 8 below. Interestingly, the $CHANGE\_SFR$ variable was unprofitable over the entire period. In contrast, both the SFR and the LN_SFR produced stable and generally upward cumulative excess returns over the sample period. This is consistent with the positive and statistically significant coefficients of the univariate analyses of SFR in Table 5.

The decline of the cumulative excess returns of the SFR and the LN_SFR in early 2021 is likely due to both the COVID-19 pandemic and the Gamestop short squeeze, as investors became increasingly cautious with regards to short selling (Demirer et al., 2019; Chohan, 2021). Nevertheless, a strategy of longing stocks with high short float ratios is likely to yield profits, as short sellers will, at some point, need to cover their short positions. This results in an increase in demand and, subsequently, an increase in prices for shorted assets (Asquith et al., 2005).

Figure 8:

Cumulative Correlation - Short Float Ratio and Excess Return

10 Recommendations for Further Research

Due to data availability constraints, this paper utilised monthly data to investigate the relationship between short interest and future excess returns in the cross-section. As stated by Brent et al., (1990), this may have limited the study, as short interest data is made available to investors both at the middle and end of each month. Schindler, (2015) went on to argue that the low days-to-cover ratio *(Table 1: 3.885)*, combined with short interest data that is at least 15 days behind, could lead to diverging results. In support of the prior argument, the non-lagged robustness test, suggested by Zhu et al., (2019), indicates that the coefficients of short interest, contained in Appendix C, were greater in magnitude with higher statistical significance.

Instead of comparing short interest data, which is released twice a month, to calendar month returns, Brent et al., (1990) suggested comparing mid-month short interest data to mid-month returns. However, following the Gamestop short squeeze documented by Chohan, (2021), the short interest reporting requirements outlined by Rule 4560 are currently facing review by FINRA, with the frequency of reporting likely to be increased. To overcome this limitation, and following Angel et al., (2003) and Au et al., (2009), utilising daily data is likely a better method to investigate the relationship between the two variables, as it is insusceptible to lagging errors. Although the majority of authors have analysed the relationship between short interest and excess returns one month forward, Callen and Fang, (2015) reported a statistically significant relationship between the two variables 12 months ahead. Hence, exploring the relationship at different lags is another consideration for further research.

Furthermore, the Fama and MacBeth (1973) regression approach investigated whether a linear relationship was present between short interest and future excess returns. However, as suggested by Au et al., (2009), the relationship between the two variables may be non-monotonic, as short interest has been found to have diminishing returns. Thus, utilising short interest as a linear investment signal may prove suboptimal. Therefore, a secondary method, such as the calendar time portfolio approach utilised by Aitken et al., (1998) and Desai et al., (2002), may have provided better insight into the relationship between short interest and future excess return, as the method is able to better detect non-monotonic relationships between variables.

11 Conclusion

This paper investigated the S&P500 Index's constituent securities to determine whether short interest affects the cross-section of excess share returns on a monthly basis between January 2015 and December 2021. The findings of this study are contrary to those of modern research, which argues for the speculative view, reporting a negative relationship between short interest and future excess return, implying that short interest acts as a bearish indicator. The Fama and MacBeth (1973) regression approach produces mixed empirical results, dependent on the manner in which short interest is defined.

The results indicate that no relationship exists between the short interest ratio, defined in relation to average daily trading volume, and future excess return, hence supporting the efficient market hypothesis. However, a robust and statistically significant positive relationship is reported for the short float ratio, defined in relation to equity float, implying that the variable serves as a bullish indicator and supporting the contrarian view. The differences in statistical significance between the two definitions of short interest indicate that the short interest ratio is not indicative of a latent demand for the shorted asset, whereas the short float ratio is seen to contain incremental predictive information with regard to firm-specific fundamentals.

While this paper contributes to the extant literature on the conflicting relationship between short interest and subsequent returns, it has further implications concerning short selling regulations as the sample period includes the COVID-19 pandemic, which caused an unprecedented worldwide decline in equity prices, together with a surge in volatility. Many market participants predicted that short selling would place further downward pressure on already declining markets; hence, the practice was temporarily banned in multiple European countries.

However, the Securities and Exchange Commission refrained from imposing restrictions, arguing that short selling is necessary to facilitate ordinary market trading. The empirical evidence of this study finds that no negative relationship exists between short interest and future excess returns in the United States, hence refuting the claim that short sellers aim to profit from extreme market downturns by exacerbating price declines. Short selling restrictions are thus deemed inadvisable, as short sellers' influence in the market is generally positive, as they reduce information asymmetry, aid in providing liquidity and improve pricing efficiency, regardless of whether the market is categorised as a bullish or bearish environment.

12 References

Ackert, L.F. and Athanassakos, G., 2005. The relationship between short interest and stock returns in the Canadian market. *Journal of Banking & Finance, 29(7), pp.1729-1749.*

Aggarwal, R., Saffi, P.A. and Sturgess, J., 2010. Does proxy voting affect the supply and/or demand for securities lending. *McDonough School of Business, Georgetown University.*

Aitken, M.J., Frino, A., McCorry, M.S. and Swan, P.L., 1998. Short sales are almost instantaneously bad news: Evidence from the Australian Stock Exchange. *The Journal of Finance, 53(6), pp.2205-2223.*

Akbas, F., Boehmer, E., Erturk, B. and Sorescu, S.M., 2013. Short interest, returns, and fundamentals. *Returns, and Fundamentals.*

Aksu, C. and Gunay, E., 1995. An empirical analysis of the causal relationship between short interest and stock prices. *Journal of Business Finance & Accounting, 22(5), pp.733-749.*

Albert Jr, R.L., Smaby, T.R. and Robison, H.D., 1997. Short selling and trading abuses on Nasdaq. *Financial Services Review, 6(1), pp.27-39.*

Anand, A. and Pathak, J., 2022. The role of Reddit in the GameStop short squeeze. *Economics Letters, 211, pp.1249.*

Angel, J.J. and McCabe, D.M., 2009. The business ethics of short selling and naked short selling. *Journal of Business Ethics, 85(1), pp.239-249.*

Angel, J.J., Christophe, S.E. and Ferri, M.G., 2003. A close look at short selling on Nasdaq. *Financial Analysts Journal, 59(6), pp.66-74.*

Arnold, T., Butler, A.W., Crack, T.F. and Zhang, Y., 2005. The information content of short interest: A natural experiment. *The Journal of Business, 78(4), pp.1307-1336.*

Asquith, P., Pathak, P.A. and Ritter, J.R., 2005. Short interest, institutional ownership, and stock returns. *Journal of Financial Economics, 78(2), pp.243-276.*

Au, A.S., Doukas, J.A. and Onayev, Z., 2009. Daily short interest, idiosyncratic risk, and stock returns. *Journal of Financial Markets, 12(2), pp.290-316.*

Autore, D.M., Boulton, T.J. and Braga-Alves, M.V., 2015. *Failures to deliver, short sale constraints, and stock overvaluation. Financial Review, 50(2), pp.143-172.*

Bailey, W. and Zheng, L., 2013. Banks, bears, and the financial crisis. *Journal of Financial Services Research, 44(1), pp.1-51.*

Barber, B.M., Lyon, J.D. and Tsai, C.L., 1999. Improved methods for tests of long-run abnormal stock returns. *The Journal of Finance, 54(1), pp.165-201.*

Baron, D.C. and McDonald, J.G., 1973. Risk and return on short positions in common stocks. *The Journal of Finance, 28(1), pp.97-107.*

Beber, A. and Pagano, M., 2013. Short-selling bans around the world: Evidence from the 2007–09 crisis. *The Journal of Finance, 68(1), pp.343-381.*

Biggs, B.M., 1966. The Short Interest—A False Proverb. *Financial Analysts Journal, 22(4), pp.111-116.*

Boehme, R.D., Danielsen, B.R. and Sorescu, S.M., 2006. Short-sale constraints, differences of opinion, and overvaluation. *Journal of Financial and Quantitative Analysis, 41(2), pp.455-487.*

Boehmer, E. and Wu, J., 2013. Short selling and the price discovery process. *The Review of Financial Studies, 26(2), pp.287-322.*

Boehmer, E., Huszar, Z.R. and Jordan, B.D., 2010. The good news in short interest. J*ournal of Financial Economics, 96(1), pp.80-97.*

Boehmer, E., Jones, C.M. and Zhang, X., 2008. Which shorts are informed? *The Journal of Finance, 63(2), pp.491-527.*

Boni, L., 2006. Strategic delivery failures in US equity markets. *Journal of Financial Markets, 9(1), pp.1-26.*

Boulton, T.J. and Braga-Alves, M.V., 2012. Naked short selling and market returns. *The Journal of Portfolio Management, 38(3), pp.133-142.*

Brent, A., Morse, D. and Stice, E.K., 1990. Short interest: Explanations and tests. *Journal of Financial and Quantitative Analysis, 25(2), pp.273-289.*

Bris, A., Goetzmann, W.N. and Zhu, N., 2007. Efficiency and the bear: Short sales and markets around the world. *The Journal of Finance, 62(3), pp.1029-1079.*

Callen, J.L. and Fang, X., 2015. Short interest and stock price crash risk. *Journal of Banking & Finance, 60, pp.181-194.*

Caster, P., and Vu, J.D., 1987. Why all the interest in short interest?. *Financial Analysts Journal, 43(4), pp.76-79.*

Chang, E.C., Cheng, J.W. and Yu, Y., 2007. Short-sales constraints and price discovery: Evidence from the Hong Kong market. *The Journal of Finance, 62(5), pp.2097-2121.*

Chang, L., Chen, H. and Zhu, Y., 2019. Short-selling constraints and corporate payout policy. *Accounting & Finance, 59(4), pp.2273-2305.* Charoenrook, A. and Daouk, H., 2009. A study of market-wide short-selling restrictions. *Available at SSRN 687562*

Chohan, U.W., 2021. Counter-hegemonic finance: The GameStop Short Squeeze. *Available at SSRN 3775127.*

Cristian, M.M. and Raisa, M.L., 2018. Short-Selling Regulation and the Development of the Stock Markets. *Ovidius University Annals, Economic Sciences Series, 18(1), pp.470-475.*

D'Avolio, G., 2002. The market for borrowing stock. *Journal of financial economics, 66(2-3), pp.271-306.*

Davis, H.A., 2012. Selected FINRA notices and disciplinary actions, June-August 2012. *Journal of Investment Compliance.*

De Marchi, N. and Harrison, P., 1994. Trading "in the Wind" and with Guile: The Troublesome Matter of the Short Selling of Shares in Seventeenth-Century Holland. *Higgling: transactors and their markets in the history of economics, 26, pp.47-53.*

Dechow, P.M., Hutton, A.P., Meulbroek, L. and Sloan, R.G., 2001. Short-sellers, fundamental analysis, and stock returns. *Journal of financial Economics, 61(1), pp.77-106.*

Demirer, R., Demos, G., Gupta, R. and Sornette, D., 2019. On the predictability of stock market bubbles: evidence from LPPLS confidence multi-scale indicators. *Quantitative Finance, 19(5), pp.843-858.*

Desai, H., Ramesh, K., Thiagarajan, S.R. and Balachandran, B.V., 2002. An investigation of the informational role of short interest in the Nasdaq market. *The Journal of Finanae, 57(5), pp.2263-2287.*

Diamond, D.W. and Verrecchia, R.E., 1987. Constraints on short-selling and asset price adjustment to private information. *Journal of Financial Economics, 18(2), pp.277-311.*

Diether, K.B., Lee, K.H. and Werner, I.M., 2009. Short-sale strategies and return predictability. *The Review of Financial Studies, 22(2), pp.575-607.*

Duffie, D., Garleanu, N. and Pedersen, L.H., 2002. Securities lending, shorting, and pricing. *Journal of Financial Economics, 66(2-3), pp.307-339.*

Dyl, E.A., 1978. Short selling and the capital gains tax. *Financial Analysts Journal, 34(2), pp.61-64.*

Engelberg, J.E., Reed, A.V. and Ringgenberg, M.C., 2012. How are shorts informed?: Short sellers, news, and information processing. *Journal of Financial Economics, 105(2), pp.260-278.*

Enriques, L. and Pagano, M., 2020. Emergency Measures for Equity Trading: The Case Against Short-Selling Bans and Stock Exchange Shutdowns. *Global Pandemic Crisis and Financial Stability, European Corporate Governance Institute-Law Working Paper, 513.*

Epstein, G., 1995. Short can be sweet. *Barron's, 75, p.32.*

Evans, R.B., Geczy, C.C., Musto, D.K. and Reed, A.V., 2009. Failure is an option: Impediments to short selling and options prices. *The Review of Financial Studies, 22(5), pp.1955-1980.*

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

Fama, E.F. and MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. *Journal of political economy, 81(3), pp.607-636.*

Fantazzini, D. and Maggi, M., 2012. Short selling in emerging markets: A comparison of market performance during the global financial crisis. *Handbook of short selling, pp.339-352.*

Field, A., 2009. Discovering Statistics using SPSS. *Sage Publications.*

Figlewski, S. and Webb, G.P., 1993. Options, short sales, and market completeness. *The Journal of Finance, 48(2), pp.761-777.*

Figlewski, S., 1981. The informational effects of restrictions on short sales: Some empirical evidence. *Journal of Financial and Quantitative Analysis, 16(4), pp.463-476.*

Fosback, N.G., 1993. Stock market logic: A sophisticated approach to profits on Wall Street. *Dearborn Trade Pub.*

Fotak, V., Raman, V. and Yadav, P.K., 2009. Naked short sellers: Angels or barbarians?. *University of Oklahoma, Price College of Business.*

Geczy, C.C., Musto, D.K. and Reed, A.V., 2002. Stocks are special too: An analysis of the equity lending market. *Journal of Financial Economics, 66(2-3), pp.241-269.*

Hanna, M., 1976. A stock price predictive model based on changes in ratios of short interest to trading volume. *Journal of Financial and Quantitative Analysis, 11(5), pp.857-872.*

Haruvy, E. and Noussair, C.N., 2006. The effect of short selling on bubbles and crashes in experimental spot asset markets. T*he Journal of Finance, 61(3), pp.1119-1157.*

Helmes, U., Henker, J. and Henker, T., 2010. The effect of the ban on short selling on market efficiency and volatility. *Proceedings of the 2011 FMA Annual Meeting (pp. 1-54). Financial Management Association.*

Henry, T.R. and Koski, J.L., 2010. Short selling around seasoned equity offerings. *The Review of Financial Studies, 23(12), pp.4389-4418.*

Huang, S., O'Hara, M. and Zhong, Z., 2021. Innovation and informed trading: Evidence from industry ETFs. *The Review of Financial Studies, 34(3), pp.1280-1316.*

Hurtado-Sanchez, L., 1978. Short interest: its influence as a stabilizer of stock returns. *Journal of Financial and quantitative analysis, 13(5), pp.965-985.*

Jaffe, J.F., 1974. Special information and insider trading. *The Journal of Business, 47(3), pp.410-428.*

Jones, C.M. and Lamont, O.A., 2002. Short-sale constraints and stock returns. *Journal of Financial Economics, 66(2-3), pp.207-239.*

Kerrigan, T.J., 1974. The short interest ratio and its component parts. *Financial analysts journal, 30(6), pp.45-49.*

Khanna, N. and Mathews, R., 2010. Bear Raids and Short Sale Bans: Is Government Intervention Justifiable?. *Journal of Finance 63 (2008): 2703-2736.*

Lasser, D.J., Wang, X. and Zhang, Y., 2010. The effect of short selling on market reactions to earnings announcements. *Contemporary Accounting Research, 27(2), pp.609-638.*

Luhr, S., 1992. Securities Lending. *International Securities Lending, pp. 1-6.* Mandelker, G., 1974. Risk and return: The case of merging firms. *Journal of financial economics, 1(4), pp.303-335.*

Mayor, T.H., 1968. Short trading activities and the price of equities: Some simulation and regression results. *Journal of Financial and Quantitative Analysis, 3(3), pp.283-298.*

Meeker, J.E., 1932. Short selling. *New York: Harper Brothers, 1932.*

Mohamad, A., Jaafar, A. and Goddard, J., 2016. Short selling and exchange-traded funds returns: evidence from the London Stock Exchange. *Applied Economics, 48(2), pp.152-164.*

Nezafat, P., Shen, T., Wang, Q. and Wu, J., 2019. Longs, shorts, and the cross-section of stock returns. *Shorts, and the Cross-Section of Stock Returns.*

Oesterle, D.A., 2006. Regulating hedge funds. *Entrepreneurial Bus. LJ, 1, pp. 1-7.*

Peters, E.E., 1996. Chaos and order in the capital markets: a new view of cycles, prices, and market volatility. *John Wiley Sons.*

Phillips, B., 2011. Options, short-sale constraints and market efficiency: A new perspective. *Journal of Banking & Finance, 35(2), pp.430-442.*

Pinches, G.E., 1970. The random walk hypothesis and technical analysis. *Financial Analysts Journal, 26(2), pp.104-110.*

Platt, H., 2004. A fuller theory of short selling. *Journal of Asset Management, 5(1), pp.49-63.*

Poitras, G., 2009. From Antwerp to Chicago: The history of exchange traded derivative security contracts. *Revue d'Histoire des Sciences Humaines, (1), pp.11-50.*

Ramsay, I., 1993. Short Selling: Further Issues. *Securities Regulation Law Journal, 21(2).*

Rapach, D.E., Ringgenberg, M.C. and Zhou, G., 2016. Short interest and aggregate stock returns. *Journal of Financial Economics, 121(1), pp.46-65.*

Reed, A.V., 2002. Costly short-selling and stock price adjustment to earnings announcements. *University of Pennsylvania.*

Reed, A.V., 2013. Short selling. *Annu. Rev. Financ. Econ., 5(1), pp.245-258.*

Renshaw, E.F., 1977. Short Selling and Financial Arbitrage. *Financial Analysts Journal, 33(1), pp.58-65.*

S&P Dow Jones Indices. 2022. S&P U.S. Indices Methodology. Available:

Schindler, D., 2015. The effect of short interest on the subsequent stock performance in NASDAQ-100 companies. *University of Twente.*

Seneca, J.J., 1967. Short interest: bearish or bullish?. *The Journal of Finance, 22(1), pp.67-70.*

Shkilko, A., Van Ness, B. and Van Ness, R., 2012. Short selling and intraday price pressures. *Financial Management, 41(2), pp.345-370.*

Siciliano, G. and Ventoruzzo, M., 2020. Banning Cassandra from the Market? An Empirical Analysis of Short-Selling Bans during the Covid-19 Crisis. *European Company and Financial Law Review, 17(3-4), pp.386-418.*

Smidt, S., 1968. A new look at the random walk hypothesis. *Journal of Financial and Quantitative Analysis, 3(3), pp.235-261.*

Smith, R.D., 1968. Short interest and stock market prices. *Financial Analysts Journal, 24(6), pp.151-154.*

Staley, K.F., 1996. The art of short selling (Vol. 4). *John Wiley & Sons.*

Stanley, C.A., 2009. The panic effect: possible unintended consequences of the temporary bans on short selling enacted during the 2008 financial crisis. *Entrepreneurial Bus. LJ, 4, pp.267-280.*

van Dillen, J.G., 1930. Isaac le Maire en de handel in actiën der Oost-Indische Compagnie. *Economisch-Historisch Jaarboek, pp.1–165*

Van Dillen, J.G., Poitras, G. and Majithia, A., 2006. Isaac Le Maire and the early trading in Dutch East India company shares. *Pioneers of Financial Economics, 1, pp.45-63.*

Vryghem, A., 2017. Analyzing the statistical significance and potential impact of abnormal returns caused by a short squeeze. *University of Twente.*

Werner, S.P., 2010. Short selling activities and convertible bond arbitrage: *Empirical evidence from the New York Stock Exchange (Vol. 75).*

Wooldridge, J.M., 2015. Introductory Econometrics: A Modern Approach. *Cengage learning.*

Woolridge, J.R. and Dickinson, A., 1994. Short selling and common stock prices. *Financial Analysts Journal, 50(1), pp.20-28.*

Zhu, Z., Duan, X. and Tu, J., 2019. The Trend in Short Selling and the Cross Section of Stock Returns. *Annals of Economics and Finance, 20(2), pp.565-586.*

13 Appendices

13.1 Appendix A: Histograms of Unbalanced Data

Figure 9: Appendix A:

Histogram of Change in SIR

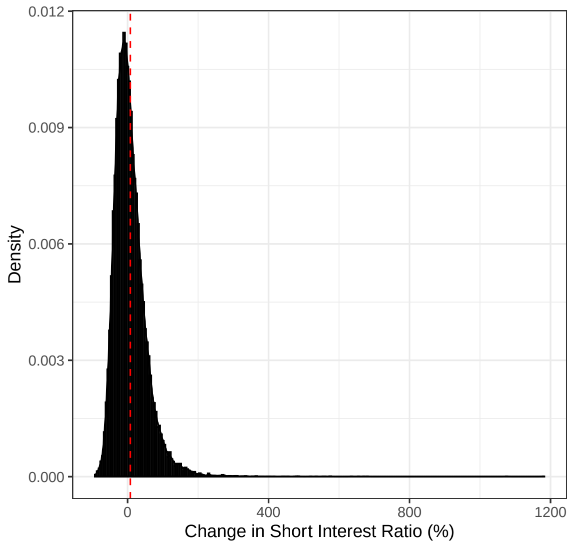


Figure 10: Appendix A:

Histogram of Change in SFR

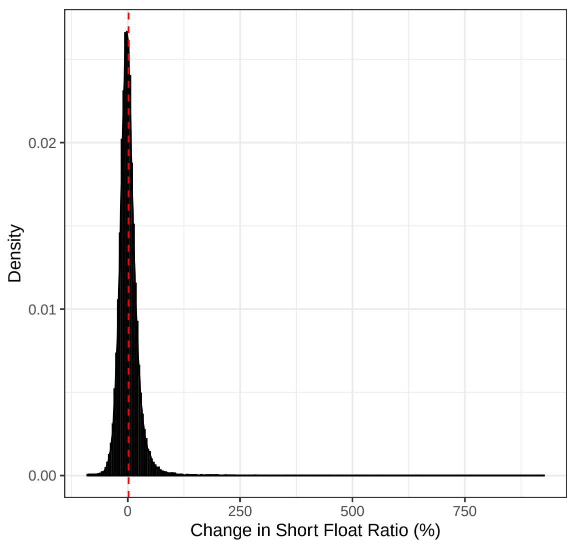


Figure 11: Appendix A:

Histogram of Earnings Yield

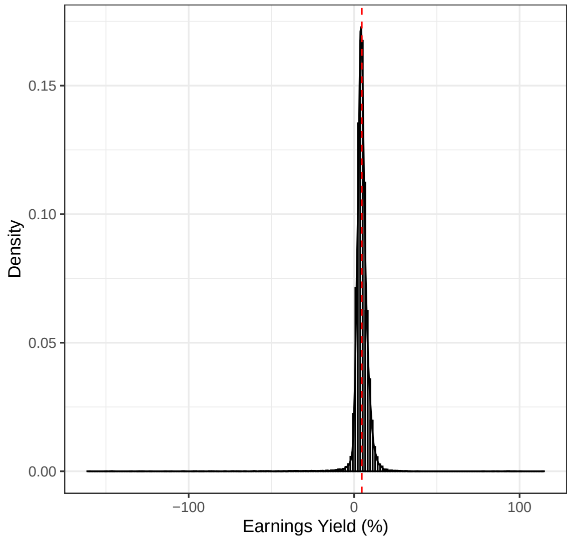


Figure 12: Appendix A:

Histogram of Price Reversal Effect

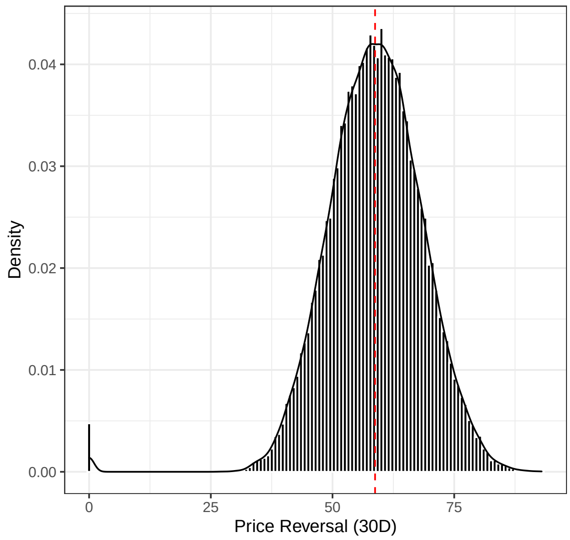


Figure 13: Appendix A:
Histogram of Momentum

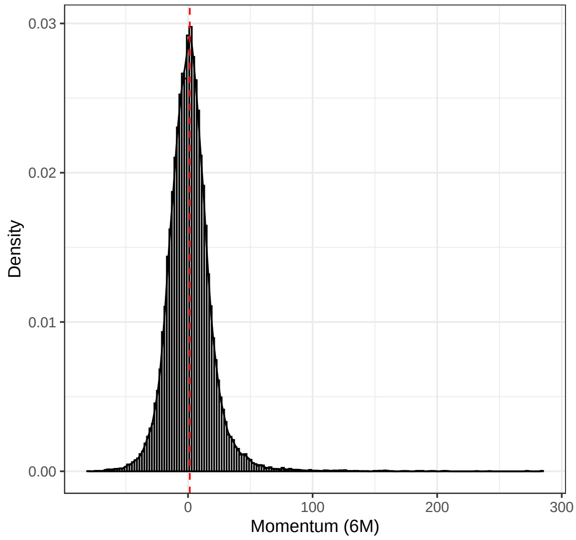


Figure 14: Appendix A:
Histogram of Market Capitalisation

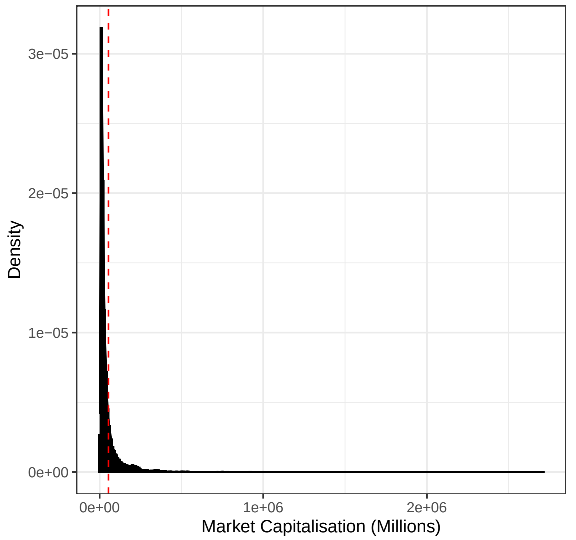


Figure 15: Appendix A:

Histogram of Natural Log of Market Capitalisation

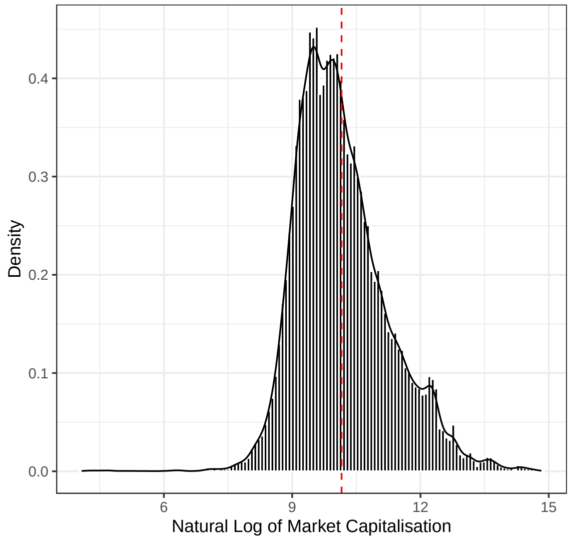
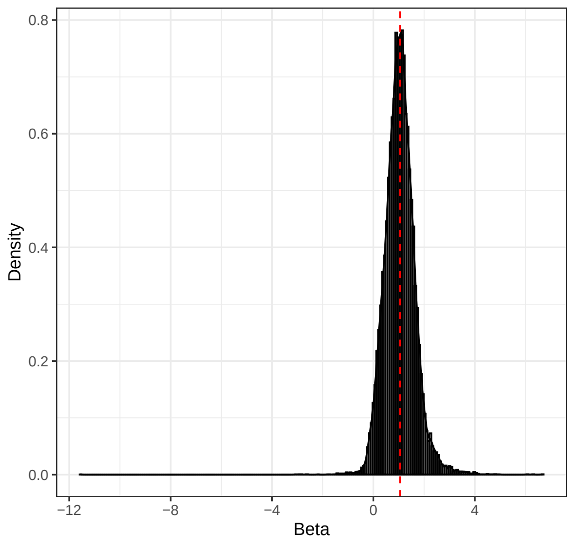


Figure 16: Appendix A:

Histogram of Beta



13.2 Appendix B: Results of Balanced Data

Table 7: Appendix B: Correlation Matrix (Balanced)

	ER	SIR	LN_SIR	CHANGE_SIR	SFR	LN_SFR	CHANGE_SFR	EY	PXREV	MOM	MARKET_CAP	LN_MARKET_CAP	BETA
ER	1												
SIR	0.001	1											
LN_SIR	-0.008	**0.913**	1										
CHANGE_SIR	-0.009	**0.223**	**0.270**	1									
SFR	0.020	**0.680**	**0.620**	0.060	1								
LN_SFR	0.019	**0.701**	**0.725**	0.088	**0.877**	1							
CHANGE_SFR	-0.011	0.052	0.082	**0.412**	**0.251**	**0.231**	1						
EY	-0.018	-0.068	-0.073	-0.009	-0.088	-0.070	0.022	1					
PXREV	0.005	-0.075	-0.076	0.000	**−0.222**	**−0.255**	-0.030	-0.189	1				
MOM	0.012	-0.012	-0.012	-0.022	-0.053	-0.041	-0.055	-0.123	**0.365**	1			
MARKET_CAP	-0.005	-0.196	**−0.254**	-0.019	**-0.217**	**−0.358**	-0.011	-0.002	0.151	0.047	1		
LN_MARKET_CAP	-0.043	**−0.311**	**−0.351**	**-0.286**	**−0.381**	**−0.535**	**-0.297**	0.077	0.199	0.037	**0.742**	1	
BETA	0.037	-0.109	-0.127	0.005	0.028	0.051	0.001	0.062	-0.198	0.004	-0.019	-0.085	1

Table 8: Appendix B: Univariate Analyses of SIR (Balanced)

	Dependent variable:		
	ER		
	(1)	(2)	(3)
SIR	0.001 (0.047)		
LN_SIR		0.083 (0.715)	
CHANGE_SIR		−0.002 (-1.615)	
Constant	1.413*** (2.621)	1.512*** (2.666)	1.364*** (2.696)
Observations	34,860	34,860	34,860
R^2	0.300	0.301	0.299

Significance Levels: *p<0.1; **p<0.05; ***p<0.01

Numbers in brackets represent two-tailed t-statistics

Table 9: Appendix B: Multiple Regression Analyses of SIR (Balanced)

	Dependent variable:		
	ER		
	(1)	(2)	(3)
SIR	0.014 (0.711)		
LN_SIR		0.019 (0.197)	
CHANGE_SIR			−0.002 (-1.583)
EY	−0.037 (-1.441)	−0.039 (-1.517)	−0.039 (-1.500)
PXREV	0.015 (0.781)	0.014 (0.744)	0.015 (0.775)
MOM	0.001 (0.267)	0.001 (0.301)	0.001 (0.221)
BETA	0.550* (1.715)	0.541* (1.703)	0.559* (1.728)
Constant	0.025 (0.020)	0.163 (0.130)	0.058 (0.049)
Observations	34,860	34,860	34,860
R^2	0.395	0.396	0.395

Significance Level: *p<0.1; **p<0.05; ***p<0.01

Numbers in brackets two-tailed represent t-statistics

Table 10: Appendix B: Univariate Analyses of SFR (Balanced)

	Dependent variable:		
	ER		
	(1)	(2)	(3)
SFR	0.050 (1.579)		
LN_SFR		0.216* (1.710)	
CHANGE_SFR			−0.004* (-1.821)
Constant	1.251*** (2.625)	1.243*** (2.670)	1.406*** (2.849)
Observations	34,860	34,860	34,860
R^2	0.308	0.307	0.299

Significance Level: *p<0.1; **p<0.05; ***p<0.01

Numbers in brackets represent two-tailed t-statistics

Table 11: Appendix B:
Multiple Regression Analyses of SFR
(Balanced)

	Dependent variable:		
	ER		
	(1)	(2)	(3)
SFR	0.043* (1.784)		
LN_SFR		0.171* (1.837)	
CHANGE_SFR			−0.003 (-1.458)
EY	−0.035 (-1.362)	−0.037 (-1.442)	−0.038 (-1.491)
PXREV	0.019 (1.000)	0.019 (0.978)	0.014 (0.721)
MOM	0.001 (0.283)	0.001 (0.224)	0.001 (0.284)
BETA	0.534* (1.656)	0.520* (1.712)	0.533* (1.640)
Constant	−0.294 (-0.294)	−0.265 (-0.216)	0.167 (0.141)
Observations	34,860	34,860	34,860
R^2	0.399	0.398	0.394

Significance Level: *p<0.1; **p<0.05; ***p<0.01

Numbers in brackets represent two-tailed t-statistics

13.3 Appendix C: Results of Non-lagged Data

Table 12: Appendix C:

Univariate Analyses of SIR

(Non-lagged)

	Dependent variable:		
	ER		
	(1)	(2)	(3)
SIR	0.076*** (2.894)		
LN_SIR		0.453*** (3.378)	
CHANGE_SIR			0.009*** (5.391)
Constant	1.018* (1.829)	0.799 (1.3608)	1.276** (2.525)
Observations	38,087	38,087	38,087
R^2	0.301	0.304	0.302

Note: *p<0.1; **p<0.05; ***p<0.01

Numbers in brackets represent two-tailed t-statistics

Table 13: Appendix C:

Multiple Regression Analyses of SIR

(Non-lagged)

	Dependent variable:		
	ER		
	(1)	(2)	(3)
SIR	0.101*** (6.055)		
LN_SIR		0.594*** (6.692)	
CHANGE_SIR			0.008*** (6.168)
EY	−0.060*** (-2.774)	−0.056*** (-2.638)	−0.067*** (-3.134)
PXREV	0.168*** (11.358)	0.169*** (11.429)	0.165*** (11.276)
MOM	−0.00002* (-1.871)	−0.00002* (-1.915)	−0.00001* (-1.755)
BETA	0.941* (1.769)	0.967* (1.824)	0.878* (1.653)
Constant	−9.678*** (-10.747)	−10.077*** (-11.086)	−9.064*** (-10.294)
Observations	38,087	38,087	38,087
R^2	0.483	0.485	0.484

Note: *p<0.1; **p<0.05; ***p<0.01

Numbers in brackets represent two-tailed t-statistics

Table 14: Appendix C:
Univariate Analyses of SFR
(Non-lagged)

	Dependent variable:		
	ER		
	(1)	(2)	(3)
SFR	0.043 (1.146)		
LN_SFR		0.285* (1.799)	
CHANGE_SFR			0.002 (0.926)
Constant	1.138** (2.439)	1.048** (2.301)	1.282** (2.546)
Observations	38,087	38,087	38,087
R^2	0.312	0.312	0.301
Note:	*p<0.1; **p<0.05; ***p<0.01		

Numbers in brackets represent two-tailed t-statistics

Table 15: Appendix C:
Multiple Regression Analyses of SFR
(Non-lagged)

	Dependent variable:		
	ER		
	(1)	(2)	(3)
SFR	0.103*** (4.289)		
LN_SFR		0.599*** (6.479)	
CHANGE_SFR			0.006*** (3.037)
EY	−0.055** (-2.550)	−0.053** (-2.509)	−0.069*** (-3.168)
PXREV	0.174*** (11.791)	0.177*** (11.939)	0.166*** (11.213)
MOM	−0.00002** (-1.964)	−0.00002** (-2.181)	−0.00001* (-1.696)
BETA	0.872* (1.649)	0.854 (1.618)	0.898* (1.698)
Constant	−9.911*** (-11.041)	−10.284*** (-11.339)	−9.070*** (-10.191)
Observations	38,087	38,087	38,087
R^2	0.488	0.488	0.483
Note:	*p<0.1; **p<0.05; ***p<0.01		

Numbers in brackets represent two-tailed t-statistics

13.4 Appendix D: RStudio Code

LOADING REQUIRED PACKAGES

Installed a package which allows for panel data econometrics in R. The Fama-Macbeth method can be run using a Mean-Grouped estimator, where we simply swap the “group” and “time” indices. Copy and paste following link into your internet browser for more details: https://www.r-bloggers.com/2012/06/fama-macbeth-and-cluster-robust-by-firm-and-time-standard-errors-in-r/ if (!require(plm)) install.packages(“plm”) library(plm)

Installed Linear, generalized linear, and nonlinear mixed models package. Allows for regressions to be run easily on subsets, in this case “Dates” if (!require(lme4)) install.packages(“lme4”) library(lme4)

Installed a package which allows for coefficient tests to be calculated if (!require(lmtest)) install.packages(“lmtest”) library(lmtest)

Installed a package which allows for Descriptive Statistics of the variables to be calculated if (!require(moments)) install.packages(“moments”) library(moments)

Installed packages which allows for graphs to be constructed if (!require(plotrix)) install.packages(“plotrix”) library(plotrix) if (!require(ggplot2)) install.packages(“ggplot2”) library(ggplot2)

Installed packages which allows for correlations to be calculated if (!require(corrplot)) install.packages(“corrplot”) library(corrplot) if (!require(Hmisc)) install.packages(“Hmisc”) library(Hmisc)

Installed a package allowing LateX code to be rendered if (!require(stargazer)) install.packages(“stargazer”) library(stargazer)

Installed a package to tidy the messy output which R produces and allows other packages to produce tables if (!require(broom)) install.packages(“broom”) library(broom)

Installed a package to allow the pipe operator (if (!require(dplyr)) install.packages(“dplyr”) library(dplyr)

IMPORTING THE DATASET

Excess Returns are calculated by taking the 30th January 2015 Total Returns (which are prices adjusted for dividends and stock-splits) divided by the 31st December 2014 Total returns and subtracting 1. The risk-free rate (1 month U.S treasury bill) is then subtracted from the monthly returns Therefore, the monthly Excess Returns are calculated. The predictor variables refer to the values at the end of the previous month. Therefore, January 2015 returns are aligned with the predictor variables values at the end of December 2014, such that predictor variable values are given at the start of the period and return is over the period. The predictor is what is known at the start of January, which allows market participants to place their bets for the month

library(readxl) cleaneddataunbalanced readexcel(“ /Documents/Masters/Dissertation - FTX5003W/ ShortInterestRegressionKyleCumming/Data/ cleaneddataunbalanced.xlsx”, coltypes = c(“date”, “text”, “numeric”, “numeric”, “numeric”, “numeric”, “numeric”, “numeric”, “numeric”, “numeric”, “numeric”, “numeric”, “numeric”, “numeric”, “numeric”, “numeric”, “numeric”, “numeric”)) View(cleaneddataunbalanced)

Attaching Dataset so that variables are easier to call
attach(cleaneddataunbalanced)

class(cleaneddataunbalancedDates)

cleaneddataunbalancedDates as.Date(cleaneddataunbalancedDates, format = “

Time Series of Short Interest Measures SIRAVERAGE cleaneddataunbalanced groupby(Dates) summarize(Average=mean(SIR))

SFRAVERAGE cleaneddataunbalanced groupby(Dates) summarize(Average=mean(SFR))

library(readxl) SHORTINTTIME readexcel(” /Documents/Masters/Dissertation - FTX5003W/ ShortInterestRegressionKyleCumming/Data/SHORTINTTIME.xlsx“, coltypes = c(”date“, ”numeric“, ”numeric“)) View(SHORTINTTIME)

SIRSFRTIME ggplot(SHORTINTTIME, aes(x = Dates)) + geomline(aes(y = (SIRTIME), colour = ”red“)) + geomline(aes(y = (SFRTIME), colour = ”green“)) + scalecoloridentity(name = ”SI Measures“, breaks = c(”green“, ”red“), labels = c(”SFR“, ”SIR“), guide = ”legend“) + theme(plot.title = elementtext(hjust = 0.5)) + theme(plot.title = elementtext(face = ”bold“))

SIRSFRTIME

SIRSFRTIMEFINAL SIRSFRTIME + theme(axis.text = elementtext(angle = 90, hjust = 0)) + xlab(”Dates“) + ylab(”Short Interest Measures“) + ggtitle(”Monthly Short Interest Measures Over Time“)

SIRSFRTIMEFINAL + theme(legend.position = c(0.08,0.9), legend.box.background = elementrect(color = ”black“, size = 1.5))

DESCRIPTIVE STATISTICS

Providing a Descriptive Summary of the variables summary(cleaneddataunbalanced)

Descriptive Statistics for LaTeX stargazer(as.data.frame(cleaneddataunbalanced)) Provides Number of observations, mean, standard deviation, min and max for each variable. The min and max have been replaced by skewness and kurtosis SKEWNESS AND KURTOSIS

Skewness: Symmetric: Values between -0.5 to 0.5 Moderated Skewed data: Values between -1 and -0.5 or between 0.5 and 1 Highly Skewed data: Values less than -1 or greater than 1

Kurtosis : Mesokurtic: This is the normal distribution Leptokurtic: This distribution has fatter tails and a sharper peak. The kurtosis is "positive" with a value greater than 3 Platykurtic: The distribution has a lower and wider peak and thinner tails. The kurtosis is "negative" with a value greater than 3

EXCESSRETURN
print(skewness(EXCESSRETURN)) 0.2477892 –¿ Symmetric
print(kurtosis(EXCESSRETURN)) 8.860115 –¿ Leptokurtic

SHORTINTERESTRATIO
print(skewness(SIR)) 2.656543 –¿ Highly Skewed
print(kurtosis(SIR)) 13.67422 –¿ Leptokurtic

LNSIR
print(skewness(LNSIR)) 0.3339438 –¿ Symmetric
print(kurtosis(LNSIR)) 3.245241 –¿ Slightly Leptokurtic

CHANGESIR
print(skewness(CHANGESIR)) 2.915986 –¿ Highly Skewed
print(kurtosis(CHANGESIR)) 34.94992 –¿ Leptokurtic

SFR
print(skewness(SFR)) 3.409304 –¿ Highly Skewed
print(kurtosis(SFR)) 20.55984 –¿ Leptokurtic

LNSFR
print(skewness(LNSFR)) 0.5692193 –¿ Moderately Skewed
print(kurtosis(LNSFR)) 2.998453 –¿ Slightly Leptokurtic

CHANGESFR
print(skewness(CHANGESFR)) 7.12169 –¿ Highly Skewed
print(kurtosis(CHANGESFR)) 144.8201 –¿ Leptokurtic

EARNYLD
print(skewness(EARNYLD)) -5.422455 –¿ Highly Skewed
print(kurtosis(EARNYLD)) 226.4209 –¿ Leptokurtic

RSI30D
print(skewness(RSI30D)) 0.07510465 –¿ Symmetric
print(kurtosis(RSI30D)) 2.821278 –¿ Mesokurtic

RELSHRPXMOM
print(skewness(RELSHRPXMOM)) 1.809544 –¿ Highly Skewed
print(kurtosis(RELSHRPXMOM)) 19.65994 –¿ Leptokurtic

MARKET CAP
print(skewness(MARKETCAP)) 9.12333 –¿ Highly Skewed
print(kurtosis(MARKETCAP)) 124.5662 –¿ Leptokurtic

LNMARKET CAP
print(skewness(LNMARKETCAP)) 0.6039767 –¿ Moderately Skewed
print(kurtosis(LNMARKETCAP)) 4.199419 –¿ Leptokurtic

BETA
print(skewness(BETA)) 0.4269882 –¿ Symmetric
print(kurtosis(BETA)) 4.487667 –¿ Leptokurtic

HISTOGRAMS

Creating Histograms of our variables in order to visualize the data and see if there are any variables which need to be logged.

FreedmanDiaconis rule for determining the appropriate bin-width
FD function(vec) (2 * IQR(vec) / length(vec)$^($1/3))

Histogram of Excess Return
FD(EXCESSRETURN) Suggested Bin-width = 0.5485713.
Rounding up to 1 to increase readability

EXCESSRETURNHISTDENSITY
ggplot(data = cleaneddataunbalanced, aes(EXCESSRETURN)) + geomhistogram(binwidth = 1, color = "■black", fill = "white", aes(y = ..density..)) + geomdensity(alpha =0.6) + geomvline(aes(xintercept=mean(EXCESSRETURN)), color = "red", linetype = "dashed") + labs(title = "Histogram of Excess Return", x = "Excess Return (themebw() + theme(plot.title = elementtext(hjust = 0.5)) + theme(plot.title = elementtext(face = "bold"))

EXCESSRETURNHISTDENSITY

Histogram of Short Interest Ratio
FD(SIR) Suggested Bin-width = 0.154774.
Keeping this value as the graph is readable

SIRHISTDENSITY
ggplot(data = cleaneddataunbalanced, aes(SIR)) + geomhistogram(binwidth = FD(SIR), color = "black", fill = "white", aes(y = ..density..)) + geomdensity(alpha =0.6) + geomvline(aes(xintercept=mean(SIR)), color = "red", linetype = "dashed") + labs(title = "Histogram of Short Interest Ratio", x = "Short Interest Ratio", y = "Density") + themebw() + theme(plot.title = elementtext(hjust = 0.5)) + theme(plot.title = elementtext(face = "bold"))

SIRHISTDENSITY

Histogram of Natural Log of Short Interest Ratio
FD(LNSIR) Suggested Bin-width = 0.0.04801496. Keeping this value as the graph is readable

LNSIRHISTDENSITY
ggplot(data = cleaneddataunbalanced, aes(LNSIR)) + geomhistogram(binwidth = FD(LNSIR), color = "white", fill = "black", aes(y = ..density..)) + geomdensity(alpha =0.6) + geomvline(aes(xintercept=mean(LNSIR)), color = "red", linetype = "dashed") + labs(title = "Histogram of Natural Log of Short Interest Ratio", x = "Natural Log of Short Interest Ratio", y = "Density") + themebw() + theme(plot.title = elementtext(hjust = 0.5)) + theme(plot.title = elementtext(face = "bold"))

LNSIRHISTDENSITY

Histogram of Change in Short Interest Ratio
FD(CHANGESIR) Suggested Bin-width = 3.127725. Rounding up to 5 to increase readability

CHANGESIRHISTDENSITY
ggplot(data = cleaneddataunbalanced, aes(CHANGESIR)) + geomhistogram(binwidth = 5, color = "black", fill = "white", aes(y = ..density..)) + geomdensity(alpha =0.6) + geomvline(aes(xintercept=mean(CHANGESIR)), color = "red", linetype = "dashed") + labs(title = "Histogram of Change in Short Interest Ratio", x = "Change in Short Interest Ratio (themebw() + theme(plot.title = elementtext(hjust = 0.5)) + theme(plot.title = elementtext(face = "bold"))

CHANGESIRHISTDENSITY

Histogram of Short Float Ratio
FD(SFR) Suggested Bin-width = 0.146071 Rounding up to 0.25 to increase readability

SFRHISTDENSITY
ggplot(data = cleaneddataunbalanced, aes(SFR)) + geomhistogram(binwidth = 0.35, color = "black", fill = "white", aes(y = ..density..)) + geomdensity(alpha =0.6) +
geomvline(aes(xintercept=mean(SFR)), color = "red", linetype = "dashed") + labs(title = "Histogram of Short Float Ratio", x = "Short Float Ratio (themebw() + theme(plot.title = elementtext(hjust = 0.5)) + theme(plot.title = elementtext(face = "bold"))

SFRHISTDENSITY

Histogram of Natural Log of Short Float Ratio
FD(LNSFR) Suggested Bin-width = 0.06272636 Keeping this value as the graph is readable

LNSFRHISTDENSITY

ggplot(data = cleaneddataunbalanced, aes(LNSFR)) + geomhistogram(binwidth = FD(LNSFR), color = “white”, fill = “black”, aes(y = ..density..)) + geomdensity(alpha =0.6) + geomvline(aes(xintercept=mean(LNSFR)), color = “red”, linetype = “dashed”) + labs(title = “Histogram of Natural Log of Short Float Ratio”, x = “Natural Log of Short Float Ratio”, y = “Density”) + themebw() + theme(plot.title = elementtext(hjust = 0.5)) + theme(plot.title = elementtext(face = “bold”))

LNSFRHISTDENSITY

Histogram of Change in Float Interest Ratio

FD(CHANGESFR) Suggested Bin-width = 1.299974 Rounding up to 4 to increase readability

CHANGESFRHISTDENSITY

ggplot(data = cleaneddataunbalanced, aes(CHANGESFR)) + geomhistogram(binwidth = 4, color = “black”, fill = “white”, aes(y = ..density..)) + geomdensity(alpha =0.6) + geomvline(aes(xintercept=mean(CHANGESFR)), color = “red”, linetype = “dashed”) + labs(title = “Histogram of Change in Short Flaot Ratio”, x = “Change in Short Float Ratio (themebw() + theme(plot.title = elementtext(hjust = 0.5)) + theme(plot.title = elementtext(face = ”bold“))

CHANGESFRHISTDENSITY

Histogram of Earnings Yield

FD(EARNYLD) Suggested Bin-width = 0.2002832 Rounding up to 1.5 to increase readability

EARNYLDHISTDENSITY

ggplot(data = cleaneddataunbalanced, aes(EARNYLD)) + geomhistogram(binwidth = 1.5, color = ”black“, fill = ”white“, aes(y = ..density..)) + geomdensity(alpha =0.6) + geomvline(aes(xintercept=mean(EARNYLD)), color = ”red“, linetype = ”dashed“) + labs(title = ”Histogram of Earnings Yield“, x = ”Earnings Yield (themebw() + theme(plot.title = elementtext(hjust = 0.5)) + theme(plot.title = elementtext(face = “bold”))

EARNYLDHISTDENSITY

Histogram of Relative Strength Indicator (30 Day)

FD(RSI30D) Suggested Bin-width = 0.7561593 Keeping this value as the graph is readable

RSI30DHISTDENSITY

ggplot(data = cleaneddataunbalanced, aes(RSI30D)) + geomhistogram(binwidth = FD(RSI30D), color = “white”, fill = “black”, aes(y = ..density..)) + geomdensity(alpha =0.6) + geomvline(aes(xintercept=mean(RSI30D)), color = “red”, linetype = “dashed”) + labs(title = “Histogram of Relative Strength Indicator (30 Day)”, x = “RSI30D”, y = “Density”) + themebw() + theme(plot.title = elementtext(hjust = 0.5)) + theme(plot.title = elementtext(face = “bold”))

RSI30DHISTDENSITY

Histogram of Relative Share Price Momentum (6 months)

FD(RELSHRPXMOM) Suggested Bin-width = 1.154185 Rounding up to 2 to increase readability

RELSHRPXMOMHISTDENSITY

ggplot(data = cleaneddataunbalanced, aes(RELSHRPXMOM)) + geomhistogram(binwidth = 2, color = “black”, fill = “white”, aes(y = ..density..)) + geomdensity(alpha =0.6) + geomvline(aes(xintercept=mean(RELSHRPXMOM)), color = “red”, linetype = “dashed”) + labs(title = “Histogram of Relative Share Price Momentum (6 months)”, x = “RELSHRPXMOM”, y = “Density”) + themebw() + theme(plot.title = elementtext(hjust = 0.5)) + theme(plot.title = elementtext(face = “bold”))

RELSHRPXMOMHISTDENSITY

Histogram of Market Capitalization

FD(MARKETCAP) Suggested Bin-width = 2113.643 Rounding to 5000 to increase readability

MARKETCAPHISTDENSITY

ggplot(data = cleaneddataunbalanced, aes(MARKETCAP)) + geomhistogram(binwidth = 15000, color = “black”, fill = “black”, aes(y = ..density..)) + geomdensity(alpha =0.6) + geomvline(aes(xintercept=mean(MARKETCAP)), color = “red”, linetype = “dashed”) + labs(title = “Histogram of Market Capitalisation”, x = “Market Capitalisation (Millions)”, y = “Density”) + themebw() + theme(plot.title = elementtext(hjust = 0.5)) + theme(plot.title = elementtext(face = “bold”))

MARKETCAPHISTDENSITY

Histogram of Natural Log of Market Capitalization

FD(LNMARKETCAP) Suggested Bin-width = 0.0823712 Keeping this value as the graph is readable

LNMARKETCAPHISTDENSITY

ggplot(data = cleaneddataunbalanced, aes(LNMARKETCAP)) + geomhistogram(binwidth = FD(LNMARKETCAP), color = “white”, fill = “black”, aes(y = ..density..)) + geomdensity(alpha =0.6) + geomvline(aes(xintercept=mean(LNMARKETCAP)), color = “red”, linetype = “dashed”) + labs(title = “Histogram of Natural Log of Market Capitalisation”, x = “Natural Log of Market Capitalisation”, y = “Density”) + themebw() + theme(plot.title = elementtext(hjust = 0.5)) + theme(plot.title = elementtext(face = “bold”))

LNMARKETCAPHISTDENSITY

Histogram of Beta’s
FD(BETA) Suggested Bin-width = 0.04336028 Keeping this value as the graph is readable

BETAHISTDENSITY
ggplot(data = cleaneddataunbalanced, aes(BETA)) + geomhistogram(binwidth = FD(BETA), color = “black”, fill = “white”, aes(y = ..density..)) + geomdensity(alpha =0.6) + geomvline(aes(xintercept=mean(BETA)), color = “red”, linetype = “dashed”) + labs(title = “Histogram of Beta’s”, x = “Beta’s”, y = “Density”) + themebw() + theme(plot.title = elementtext(hjust = 0.5)) + theme(plot.title = elementtext(face = “bold”))

BETAHISTDENSITY

CORRELATION MATRIX

Correlation Matrix Cleaning the Dataset so that only the variables of interest remain cleaneddataunbalancednumeric cleaneddataunbalanced[, c(6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18)]

SIREXRET cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(SIR, EXCESSRETURN))
mean(SIREXRETCorrelation)

LNSIREXRET cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(LNSIR, EXCESSRETURN))
mean(LNSIREXRETCorrelation)

CHANGESIREXRET cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(CHANGESIR, EXCESSRETURN))
mean(CHANGESIREXRETCorrelation)

SFREXRET cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(SFR, EXCESSRETURN))
mean(SFREXRETCorrelation)

LNSFREXRET cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(LNSFR, EXCESSRETURN))
mean(LNSFREXRETCorrelation)

CHANGESFREXRET cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(CHANGESFR, EXCESSRETURN))
mean(CHANGESFREXRETCorrelation)

EARNYLDEXRET cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(EARNYLD, EXCESSRETURN))
mean(EARNYLDEXRETCorrelation)

RSI30DEXRET cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(RSI30D, EXCESSRETURN))
mean(RSI30DEXRETCorrelation)

RELSHREXRET cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(RELSHRPXMOM, EXCESSRETURN))
mean(RELSHREXRETCorrelation)

MARKCAPEXRET cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(MARKETCAP, EXCESSRETURN))
mean(MARKCAPEXRETCorrelation)

LNMARKCAPEXRET cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(LNMARKETCAP, EXCESSRETURN))
mean(LNMARKCAPEXRETCorrelation)

BETAEXRET cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(BETA, EXCESSRETURN))
mean(BETAEXRETCorrelation)

LNSIRSIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(LNSIR, SIR))
mean(LNSIRSIRCorrelation)

CHANGESIRSIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(CHANGESIR, SIR))
mean(CHANGESIRSIRCorrelation)

SFRSIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(SFR, SIR))
mean(SFRSIRCorrelation)

LNSFRSIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(LNSFR, SIR))
mean(LNSFRSIRCorrelation)

CHANGESFRSIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(CHANGESFR, SIR))
mean(CHANGESFRSIRCorrelation)

EARNYLDSIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(EARNYLD, SIR))
mean(EARNYLDSIRCorrelation)

RSI30DSIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(RSI30D, SIR))
mean(RSI30DSIRCorrelation)

RELSHRSIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(RELSHRPXMOM, SIR))
mean(RELSHRSIRCorrelation)

MARKCAPSIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(MARKETCAP, SIR))
mean(MARKCAPSIRCorrelation)

LNMARKCAPSIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(LNMARKETCAP, SIR))
mean(LNMARKCAPSIRCorrelation)

BETASIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(BETA, SIR))
mean(BETASIRCorrelation)

CHANGESIRLNSIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(CHANGESIR, LNSIR))
mean(CHANGESIRLNSIRCorrelation)

SFRLNSIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(SFR, LNSIR))
mean(SFRLNSIRCorrelation)

LNSFRLNSIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(LNSFR, LNSIR))
mean(LNSFRLNSIRCorrelation)

CHANGESFRLNSIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(CHANGESFR, LNSIR))
mean(CHANGESFRLNSIRCorrelation)

EARNYLDLNSIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(EARNYLD, LNSIR))
mean(EARNYLDLNSIRCorrelation)

RSI30DLNSIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(RSI30D, LNSIR))
mean(RSI30DLNSIRCorrelation)

RELSHRLNSIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(RELSHRPXMOM, LNSIR))
mean(RELSHRLNSIRCorrelation)

MARKCAPLNSIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(MARKETCAP, LNSIR))
mean(MARKCAPLNSIRCorrelation)

LNMARKCAPLNSIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(LNMARKETCAP, LNSIR))
mean(LNMARKCAPLNSIRCorrelation)

BETALNSIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(BETA, LNSIR))
mean(BETALNSIRCorrelation)

SFRCHANGESIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(SFR, CHANGESIR))
mean(SFRCHANGESIRCorrelation)

LNSFRCHANGESIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(LNSFR, CHANGESIR))
mean(LNSFRCHANGESIRCorrelation)

CHANGESFRCHANGESIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(CHANGESFR, CHANGESIR))
mean(CHANGESFRCHANGESIRCorrelation)

EARNYLDCHANGESIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(EARNYLD, CHANGESIR))
mean(EARNYLDCHANGESIRCorrelation)

RSI30DCHANGESIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(RSI30D, CHANGESIR))
mean(RSI30DCHANGESIRCorrelation)

RELSHRCHANGESIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(RELSHRPXMOM, CHANGESIR))
mean(RELSHRCHANGESIRCorrelation)

MARKCAPCHANGESIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(MARKETCAP, CHANGESIR))
mean(MARKCAPCHANGESIRCorrelation)

LNMARKCAPCHANGESIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(LNMARKETCAP, CHANGESIR))
mean(LNMARKCAPCHANGESIRCorrelation)

BETACHANGESIR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(BETA, CHANGESIR))
mean(BETACHANGESIRCorrelation)

LNSFRSFR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(LNSFR, SFR))
mean(LNSFRSFRCorrelation)

CHANGESFRSFR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(CHANGESFR, SFR))
mean(CHANGESFRSFRCorrelation)

EARNYLDSFR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(EARNYLD, SFR))
mean(EARNYLDSFRCorrelation)

RSI30DSFR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(RSI30D, SFR))
mean(RSI30DSFRCorrelation)

RELSHRSFR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(RELSHRPXMOM, SFR))
mean(RELSHRSFRCorrelation)

MARKCAPSFR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(MARKETCAP, SFR))
mean(MARKCAPSFRCorrelation)

LNMARKCAPSFR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(LNMARKETCAP, SFR))
mean(LNMARKCAPSFRCorrelation)

BETASFR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(BETA, SFR))
mean(BETASFRCorrelation)

CHANGESFRLNSFR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(CHANGESFR, LNSFR))
mean(CHANGESFRLNSFRCorrelation)

EARNYLDLNSFR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(EARNYLD, LNSFR))
mean(EARNYLDLNSFRCorrelation)

RSI30DLNSFR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(RSI30D, LNSFR))
mean(RSI30DLNSFRCorrelation)

RELSHRLNSFR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(RELSHRPXMOM, LNSFR))
mean(RELSHRLNSFRCorrelation)

MARKCAPLNSFR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(MARKETCAP, LNSFR))
mean(MARKCAPLNSFRCorrelation)

LNMARKCAPLNSFR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(LNMARKETCAP, LNSFR))
mean(LNMARKCAPLNSFRCorrelation)

BETALNSFR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(BETA, LNSFR))
mean(BETALNSFRCorrelation)

EARNYLDCHANGESFR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(EARNYLD, CHANGESFR))
mean(EARNYLDCHANGESFRCorrelation)

RSI30DCHANGESFR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(RSI30D, CHANGESFR))
mean(RSI30DCHANGESFRCorrelation)

RELSHRCHANGESFR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(RELSHRPXMOM, CHANGESFR))
mean(RELSHRCHANGESFRCorrelation)

MARKCAPCHANGESFR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(MARKETCAP, CHANGESFR))
mean(MARKCAPCHANGESFRCorrelation)

LNMARKCAPCHANGESFR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(LNMARKETCAP, CHANGESFR))
mean(LNMARKCAPCHANGESFRCorrelation)

BETACHANGESFR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(BETA, CHANGESFR))
mean(BETACHANGESFRCorrelation)

RSI30DEARNYLD cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(RSI30D, EARNYLD))
mean(RSI30DEARNYLDCorrelation)

RELSHREARNYLD cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(RELSHRPXMOM, EARNYLD))
mean(RELSHREARNYLDCorrelation)

MARKCAPEARNYLD cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(MARKETCAP, EARNYLD))
mean(MARKCAPEARNYLDCorrelation)

LNMARKCAPEARNYLD cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(LNMARKETCAP, EARNYLD))
mean(LNMARKCAPEARNYLDCorrelation)

BETAEARNYLD cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(BETA, EARNYLD))
mean(BETAEARNYLDCorrelation)

RELSHRRSI30D cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(RELSHRPXMOM, RSI30D))
mean(RELSHRRSI30DCorrelation)

MARKCAPRSI30D cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(MARKETCAP, RSI30D))
mean(MARKCAPRSI30DCorrelation)

LNMARKCAPRSI30D cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(LNMARKETCAP, RSI30D))
mean(LNMARKCAPRSI30DCorrelation)

BETARSI30D cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(BETA, RSI30D))
mean(BETARSI30DCorrelation)

MARKCAPRELSHR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(MARKETCAP, RELSHRPXMOM))
mean(MARKCAPRELSHRCorrelation)

LNMARKCAPRELSHR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(LNMARKETCAP, RELSHRPXMOM))
mean(LNMARKCAPRELSHRCorrelation)

BETARELSHR cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(BETA, RELSHRPXMOM))
mean(BETARELSHRCorrelation)

LNMARKCAPMARKCAP cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(LNMARKETCAP, MARKETCAP))
mean(LNMARKCAPMARKCAPCorrelation)

BETAMARKCAP cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(BETA, MARKETCAP))
mean(BETAMARKCAPCorrelation)

BETALNMARKCAP cleaneddataunbalanced groupby(Dates) summarize(Correlation=cor(BETA, LNMARKETCAP))
mean(BETALNMARKCAPCorrelation)

Using the stargazer package to produce the LateX code for the correlation matrix stargazer(correlation, header=FALSE, to get rid of r package output text

single.row = TRUE, to put coefficients and standard errors on same line

no.space = TRUE, to remove the spaces after each line of coefficients

column.sep.width = “3pt”, to reduce column width

font.size = “small” to make font size smaller))

The above formatting is kept but the numerical values are replacing with the mean correlations as calculated above, as the built-in correlation function in R returns the time-series correlation, not the cross-sectional correlation as calculated above.

PLOTTING PAYOFFS TO SHORT INTEREST MEASURES OVER TIME

library(readxl) SIRMEASURESEXCESSRETURNCORR readexcel(“ /Documents/Masters/Dissertation - FTX5003W/ShortInterestRegressionKyleCumming /Data/SIRMEASURESEXCESSRETURNCORR.xlsx”, coltypes = c(“date”, “numeric”, “numeric”, “numeric”)) View(SIRMEASURESEXCESSRETURNCORR)

SIRMEASURESEXCESSRETURNCORRDates as.Date(SIRMEASURESEXCESSRETURNCORRDates, format = “

SIRMEASURECORR ggplot(SIRMEASURESEXCESSRETURNCORR, aes(x = Dates)) + geomline(aes(y = cumsum(SIRCORRELATION), colour = ”red“)) + geomline(aes(y = cumsum(LNSIRCORRELATION), colour = ”green“)) + geomline(aes(y = cumsum(CHANGESIRCORRELATION), colour = ”blue“)) + scalecoloridentity(breaks = c(”blue“, ”green“, ”red“), labels = c(”CHANGE SIR“, ”LN SIR“, ”SIR“), guide = ”legend“) + theme(plot.title = elementtext(hjust = 0.5)) + theme(plot.title = elementtext(face = ”bold“))

SIRMEASURECORR

SIRCUMULATIVEGRAPH SIRMEASURECORR + theme(axis.text = elementtext(angle = 90, hjust = 0)) + xlab(”Dates“) + ylab(”Payoff to Factor“) + ggtitle(”Cumulative Correlation between Short Interest Ratio and Excess Return“) + scalexdate(limits = as.Date(c(”2014-12-31“, ”2021-12-31“)))

SIRCUMULATIVEGRAPH + theme(legend.key.size = unit(0.4, ”cm“), legend.title = elementblank(), legend.position = c(0.09,0.925), legend.box.background = elementrect(color = ”black“, size = 1.5))

library(readxl) SFRMEASURESEXCESSRETURNCORR readexcel(” /Documents/Masters/Dissertation - FTX5003W/ShortInterestRegressionKyleCumming/ Data/SFRMEASURESEXCESSRETURNCORR.xlsx“, coltypes = c(”date“, ”numeric“, ”numeric“, ”numeric“)) View(SFRMEASURESEXCESSRETURNCORR)

SFRMEASURESEXCESSRETURNCORRDates as.Date(SFRMEASURESEXCESSRETURNCORRDates, format = ”

SFRMEASURECORR ggplot(SFRMEASURESEXCESSRETURNCORR, aes(x = Dates)) + geomline(aes(y = cumsum(SFRCORRELATION), colour = “red”)) + geomline(aes(y = cumsum(LNSFRCORRELATION), colour = “green”)) + geomline(aes(y = cumsum(CHANGESFRCORRELATION), colour = “blue”)) + scalecoloridentity(name = “SFR Measures”, breaks = c(“blue”, “green”, “red”), labels = c(“CHANGE SFR”, “LN SFR”, “SFR”), guide = “legend”) + theme(plot.title = elementtext(hjust = 0.5)) + theme(plot.title = elementtext(face = “bold”))

SFRMEASURECORR

SFRCUMULATIVEGRAPH SFRMEASURECORR + theme(axis.text = elementtext(angle = 90, hjust = 0)) + xlab(“Dates”) + ylab(“Payoff to Factor”) + ggtitle(“Cumulative Correlation between Short Float Measures and Excess Return”) + scalexdate(limits = as.Date(c(“2014-12-31”, “2021-12-31”)))

SFRCUMULATIVEGRAPH + theme(legend.key.size = unit(0.4, “cm”), legend.title = elementblank(), legend.position = c(0.09,0.925), legend.box.background = elementrect(color = “black”, size = 1.5))

Using the stargazer package to produce the LateX code for the correlation matrix stargazer(correlation, header=FALSE, to get rid of r package output text

single.row = TRUE, to put coefficients and standard errors on same line

no.space = TRUE, to remove the spaces after each line of coefficients

column.sep.width = “3pt”, to reduce column width

font.size = “small” to make font size smaller))

Creating a correlation plot to add a graphical element corrplot(correlation, tl.col = “black”)

RUNNING REGRESSIONS USING THE LME4 PACKAGE, WHERE THE DATA CAN BE GROUPED BY SUBSETS, IN THIS CASE DATES

SIR Variable Alone SIR1mlag lmList(EXCESSRETURN SIR — Dates, data = cleaneddataunbalanced)

Extracting just the coefficients from the regression, as Std Errors from Regression can be ignored in line with prior research SIR1mlagcoef coef(SIR1mlag)

plot(SIR1mlagcoefSIR)

Finding the Standard Error on the coeficients std.error(SIR1mlagcoef)

Performing a t.test on the coeficients t.test(SIR1mlagcoef‘(Intercept)‘)
t.test(SIR1mlagcoefSIR)

SIR Variable as a model SIR1mlagmodel lmList(EXCESSRETURN SIR + EARNYLD + RSI30D + RELSHRPXMOM + BETA — Dates, data = cleaneddataunbalanced)

Extracting just the coefficients from the regression, as Std Errors from Regression can be ignored in line with prior research
SIR1mlagmodelcoef coef(SIR1mlagmodel)

Finding the Standard Errors on the coeficients
std.error(SIR1mlagmodelcoef)

Performing a t.test on the coeficients
t.test(SIR1mlagmodelcoef‘(Intercept)‘)
t.test(SIR1mlagmodelcoefSIR)
t.test(SIR1mlagmodelcoefEARNYLD)
t.test(SIR1mlagmodelcoefRSI30D)
t.test(SIR1mlagmodelcoefRELSHRPXMOM)
t.test(SIR1mlagmodelcoefBETA)

LNSHORTINTERESTRATIO Alone
LNSIR1mlag lmList(EXCESSRETURN LNSIR — Dates, data = cleaneddataunbalanced)

Extracting just the coefficients from the regression, as Std Errors from Regression can be ignored in line with prior research
LNSIR1mlagcoef coef(LNSIR1mlag)

Finding the Standard Error on the coeficients
std.error(LNSIR1mlagcoef)

Performing a t.test on the coeficients
t.test(LNSIR1mlagcoef‘(Intercept)‘)
t.test(LNSIR1mlagcoefLNSIR)

LNSHORTINTERESTRATIO Variable as a model
LNSIR1mlagmodel lmList(EXCESSRETURN LNSIR + EARNYLD + RSI30D + RELSHRPXMOM + BETA — Dates, data = cleaneddataunbalanced)

Extracting just the coefficients from the regression, as Std Errors from Regression can be ignored in line with prior research
LNSIR1mlagmodelcoef coef(LNSIR1mlagmodel)

Finding the Standard Error on the coefficients
std.error(LNSIR1mlagmodelcoef)

Performing a t.test on the coefficients
t.test(LNSIR1mlagmodelcoef‘(Intercept)‘)
t.test(LNSIR1mlagmodelcoefLNSIR)
t.test(LNSIR1mlagmodelcoefEARNYLD)
t.test(LNSIR1mlagmodelcoefRSI30D)
t.test(LNSIR1mlagmodelcoefRELSHRPXMOM)
t.test(LNSIR1mlagmodelcoefBETA)

CHANGESIR Variable Alone
CHANGESIR1mlag lmList(EXCESSRETURN CHANGESIR — Dates, data = cleaneddataunbalanced)

Extracting just the coefficients from the regression, as Std Errors from Regression can be ignored in line with prior research
CHANGESIR1mlagcoef coef(CHANGESIR1mlag)

Finding the Standard Error on the coefficients
std.error(CHANGESIR1mlagcoef)

Performing a t.test on the coefficients
t.test(CHANGESIR1mlagcoef‘(Intercept)‘)
t.test(CHANGESIR1mlagcoefCHANGESIR)

CHANGESIR Variable as a model
CHANGESIR1mlagmodel lmList(EXCESSRETURN CHANGESIR + EARNYLD + RSI30D + RELSHRPXMOM + BETA — Dates, data = cleaneddataunbalanced)

Extracting just the coefficients from the regression, as Std Errors from Regression can be ignored in line with prior research
CHANGESIR1mlagmodelcoef coef(CHANGESIR1mlagmodel)

Finding the Standard Error on the coefficients
std.error(CHANGESIR1mlagmodelcoef)

Performing a t.test on the coefficients
t.test(CHANGESIR1mlagmodelcoef‘(Intercept)‘)
t.test(CHANGESIR1mlagmodelcoefCHANGESIR)
t.test(CHANGESIR1mlagmodelcoefEARNYLD)
t.test(CHANGESIR1mlagmodelcoefRSI30D)
t.test(CHANGESIR1mlagmodelcoefRELSHRPXMOM)
t.test(CHANGESIR1mlagmodelcoefBETA)

SIPERCEQUITYFLOAT Variable Alone
SIPERCEQUITYFLOAT1mlag lmList(EXCESSRETURN SFR — Dates, data = cleaneddataunbalanced)

Extracting just the coefficients from the regression, as Std Errors from Regression can be ignored in line with prior research
SIPERCEQUITYFLOAT1mlagcoef coef(SIPERCEQUITYFLOAT1mlag)

Finding the Standard Error on the coeficients
std.error(SIPERCEQUITYFLOAT1mlagcoef)

Performing a t.test on the coeficients
t.test(SIPERCEQUITYFLOAT1mlagcoef‘(Intercept)‘)
t.test(SIPERCEQUITYFLOAT1mlagcoefSFR)

SIPERCEQUITYFLOAT Variable as a model
SIPERCEQUITYFLOAT1mlagmodel lmList(EXCESSRETURN SFR + EARNYLD + RSI30D + RELSHRPXMOM + BETA — Dates, data = cleaneddataunbalanced)

Extracting just the coefficients from the regression, as Std Errors from Regression can be ignored in line with prior research
SIPERCEQUITYFLOAT1mlagmodelcoef coef(SIPERCEQUITYFLOAT1mlagmodel)

Finding the Standard Errors on the coeficients
std.error(SIPERCEQUITYFLOAT1mlagmodelcoef)

Performing a t.test on the coeficients
t.test(SIPERCEQUITYFLOAT1mlagmodelcoef‘(Intercept)‘)
t.test(SIPERCEQUITYFLOAT1mlagmodelcoefSFR)
t.test(SIPERCEQUITYFLOAT1mlagmodelcoefEARNYLD)
t.test(SIPERCEQUITYFLOAT1mlagmodelcoefRSI30D)
t.test(SIPERCEQUITYFLOAT1mlagmodelcoefRELSHRPXMOM)
t.test(SIPERCEQUITYFLOAT1mlagmodelcoefBETA)

LNSFR Alone
LNSFR1mlag lmList(EXCESSRETURN LNSFR — Dates, data = cleaneddataunbalanced)

Extracting just the coefficients from the regression, as Std Errors from Regression can be ignored in line with prior research
LNSFR1mlagcoef coef(LNSFR1mlag)

Finding the Standard Error on the coeficients
std.error(LNSFR1mlagcoef)

Performing a t.test on the coeficients
t.test(LNSFR1mlagcoef'(Intercept)')
t.test(LNSFR1mlagcoefLNSFR)

LNSFR Variable as a model
LNSFR1mlagmodel lmList(EXCESSRETURN LNSFR + EARNYLD + RSI30D + RELSHRPXMOM + BETA — Dates, data = cleaneddataunbalanced)

Extracting just the coefficients from the regression, as Std Errors from Regression can be ignored in line with prior research
LNSFR1mlagmodelcoef
coef(LNSFR1mlagmodel)

Finding the Standard Error on the coefficients
std.error(LNSFR1mlagmodelcoef)

Performing a t.test on the coefficients
t.test(LNSFR1mlagmodelcoef'(Intercept)')
t.test(LNSFR1mlagmodelcoefLNSFR)
t.test(LNSFR1mlagmodelcoefEARNYLD)
t.test(LNSFR1mlagmodelcoefRSI30D)
t.test(LNSFR1mlagmodelcoefRELSHRPXMOM)
t.test(LNSFR1mlagmodelcoefBETA)

CHANGESIR Variable Alone
CHANGESFR1mlag lmList(EXCESSRETURN CHANGESFR — Dates, data = cleaneddataunbalanced)

Extracting just the coefficients from the regression, as Std Errors from Regression can be ignored in line with prior research
CHANGESFR1mlagcoef coef(CHANGESFR1mlag)

Finding the Standard Error on the coefficients
std.error(CHANGESFR1mlagcoef)

Performing a t.test on the coefficients
t.test(CHANGESFR1mlagcoef'(Intercept)')
t.test(CHANGESFR1mlagcoefCHANGESFR)

CHANGESIR Variable as a model

CHANGESFR1mlagmodel lmList(EXCESSRETURN CHANGESFR + EARNYLD + RSI30D + RELSHRPXMOM + BETA — Dates, data = cleaneddataunbalanced)

Extracting just the coefficients from the regression, as Std Errors from Regression can be ignored in line with prior research
CHANGESFR1mlagmodelcoef coef(CHANGESFR1mlagmodel)

Finding the Standard Error on the coefficients std.error(CHANGESFR1mlagmodelcoef)

Performing a t.test on the coefficients
t.test(CHANGESFR1mlagmodelcoef'(Intercept)')
t.test(CHANGESFR1mlagmodelcoefCHANGESFR)
t.test(CHANGESFR1mlagmodelcoefEARNYLD)
t.test(CHANGESFR1mlagmodelcoefRSI30D)
t.test(CHANGESFR1mlagmodelcoefRELSHRPXMOM)
t.test(CHANGESFR1mlagmodelcoefBETA)

USING THE MEANGROUPED MODEL AS A SAFETY TEST FOR VARIABLES USING THE FAMA-MACBETH METHOD, BY SWAPPING THE DATES AND FIRM WHEN INDEXING AND USING THE CALCULATED RETURNS

SIR Variable Alone FM
SIRfm1mlagret pmg(EXCESSRETURN SIR, cleaneddataunbalanced, index = c("Dates", "Firm"))

Summary of results
summary(SIRfm1mlagret)

SIR Variable as a model FM
SIRfm1mlagmodelret pmg(EXCESSRETURN SIR + EARNYLD + RSI30D + RELSHRPXMOM + BETA, cleaneddataunbalanced, index = c("Dates", "Firm"))

Summary of results
summary(SIRfm1mlagmodelret)

LNSHORTINTERESTRATIO Alone FM
LNSHORTINTERESTRATIOfm1mlagret pmg(EXCESSRETURN LNSIR, cleaneddataunbalanced, index = c("Dates", "Firm"))

Summary of results
summary(LNSHORTINTERESTRATIOfm1mlagret)

LNSHORTINTERESTRATIO Variable as a model FM
LNSHORTINTERESTRATIOfm1mlagmodelret pmg(EXCESSRETURN LNSIR + EARNYLD + RSI30D + RELSHRPXMOM + BETA, cleaneddataunbalanced, index = c("Dates", "Firm"))

Summary of results
summary(LNSHORTINTERESTRATIOfm1mlagmodelret)

CHANGESIR Variable Alone FM
CHANGESIRfm1mlagret pmg(EXCESSRETURN CHANGESIR, cleaneddataunbalanced, index = c("Dates", "Firm"))

Summary of results
summary(CHANGESIRfm1mlagret)

CHANGESIR Variable as a model FM
CHANGESIRfm1mlagmodelret pmg(EXCESSRETURN CHANGESIR + EARNYLD + RSI30D + RELSHRPXMOM + BETA, cleaneddataunbalanced, index = c("Dates", "Firm"))

Summary of results
summary(CHANGESIRfm1mlagmodelret)

Generating LateX code using Stargazer for simple regressions: stargazer(SIRfm1mlagret, LNSHORTINTERESTRATIOfm1mlagret, CHANGESIRfm1mlagret, header=FALSE, to get rid of r package output text single.row = TRUE, to put coefficients and standard errors on same line

no.space = TRUE, to remove the spaces after each line of coefficients

column.sep.width = "3pt", to reduce column width

font.size = "small" to make font size smaller))

Generating LateX code using Stargazer for multiple regressions:
stargazer(SIRfm1mlagmodelret, LNSHORTINTERESTRATIOfm1mlagmodelret, CHANGESIRfm1mlagmodelret, header=FALSE, to get rid of r package output text

single.row = TRUE, to put coefficients and standard errors on same line

no.space = TRUE, to remove the spaces after each line of coefficients

column.sep.width = "3pt", to reduce column width

font.size = "small" to make font size smaller))

SIPERCENTEQUITYFLOAT Variable Alone FM
SIPERCENTEQUITYFLOATRfm1mlagret pmg(EXCESSRETURN SFR, cleaneddataunbalanced, index = c("Dates", "Firm"))

Summary of results
summary(SIPERCENTEQUITYFLOATRfm1mlagret)

SIPERCENTEQUITYFLOAT Variable as a model FM
SIPERCENTEQUITYFLOATfm1mlagmodelret pmg(EXCESSRETURN SFR + EARNYLD + RSI30D + RELSHRPXMOM + BETA, cleaneddataunbalanced, index = c("Dates", "Firm"))

Summary of results
summary(SIPERCENTEQUITYFLOATfm1mlagmodelret)

LNSIPERCENTEQUITYFLOAT Alone FM
LNSIPERCENTEQUITYFLOATfm1mlagret pmg(EXCESSRETURN LNSFR, cleaneddataunbalanced, index = c("Dates", "Firm"))

Summary of results
summary(LNSIPERCENTEQUITYFLOATfm1mlagret)

LNSIPERCENTEQUITYFLOAT Variable as a model FM
LNSIPERCENTEQUITYFLOATfm1mlagmodelret pmg(EXCESSRETURN LNSFR + EARNYLD + RSI30D + RELSHRPXMOM + BETA, cleaneddataunbalanced, index = c("Dates", "Firm"))

Summary of results
summary(LNSIPERCENTEQUITYFLOATfm1mlagmodelret)

CHANGESIPERCENTEQUITYFLOAT Variable Alone FM
CHANGESIPERCENTEQUITYFLOATfm1mlagret pmg(EXCESSRETURN CHANGESFR, cleaneddataunbalanced, index = c("Dates", "Firm"))

Summary of results
summary(CHANGESIPERCENTEQUITYFLOATfm1mlagret)

CHANGESIPERCENTEQUITYFLOAT Variable as a model FM
CHANGESIPERCENTEQUITYFLOATfm1mlagmodelret pmg(EXCESSRETURN CHANGESFR + EARNYLD + RSI30D + RELSHRPXMOM + BETA, cleaneddataunbalanced, index = c("Dates", "Firm"))

Summary of results
summary(CHANGESIPERCENTEQUITYFLOATfm1mlagmodelret)

Generating LateX code using Stargazer for simple regressions:
stargazer(SIPERCENTEQUITYFLOATRfm1mlagret, LNSIPERCENTEQUITYFLOATfm1mlagret, CHANGESIPERCENTEQUITYFLOATfm1mlagret, header=FALSE, to get rid of r package output text

single.row = TRUE, to put coefficients and standard errors on same line

no.space = TRUE, to remove the spaces after each line of coefficients

column.sep.width = “3pt”, to reduce column width

font.size = “small” to make font size smaller))

Generating LateX code using Stargazer for multiple regressions:
stargazer(SIPERCENTEQUITYFLOATfm1mlagmodelret, LNSIPERCENTEQUITYFLOATfm1mlagmodelret, CHANGESIPERCENTEQUITYFLOATfm1mlagmodelret, header=FALSE, to get rid of r package output text

single.row = TRUE, to put coefficients and standard errors on same line

no.space = TRUE, to remove the spaces after each line of coefficients

column.sep.width = “3pt”, to reduce column width

font.size = “small” to make font size smaller))