Surprise in Short Interest *

Pavel Lesnevski[1] Esad Smajlbegovic[2]

Abstract

This paper proposes a simple and intuitive measure of *informed* short selling: *surprise in short interest*. The measure accounts for important cross-sectional, distributional differences in short selling and provides a number of novel insights. We document a significant price drift after announcements of unexpected short-selling activity. In particular, we find that stocks with a positive surprise in short interest significantly underperform stocks with a negative surprise in short interest. The resulting return spread originates from both positive and negative surprises and it is not explained by standard stock characteristics, the level of short selling or short-sale constraints. Consistent with the notion of short sellers' informed trading on mispricing, the surprise in short interest also predicts future surprises in company fundamentals. Lastly, in line with Shleifer and Vishny's (1997) limits-to-arbitrage argument, the return predictability is stronger among illiquid and volatile stocks. Overall, our results suggest that the market does not efficiently price the information from short sale reports.

Keywords: informed short selling, fundamentals, mispricing
JEL: G12, G14, G23

*We are grateful to Vikas Agarwal, Antje Berndt (discussant), Egle Karmaziene (discussant), Tzuo-Hann Law (discussant), Stefan Ruenzi, Daniel Schmidt, Michael Ungeheuer and participants at 24th Annual Meeting of the German Finance Association, 20th Annual Conference of the Swiss Society for Financial Market Research, 16th Cologne Colloquium on Financial Markets, 2017 European Joint Conference of Academy of Entrepreneurial Finance and Academy of Behavioral Finance and Economics, and the seminar at the University of Mannheim for their helpful comments and suggestions. We retain responsibility for all remaining errors.

[1]University of Mannheim, Chair of International Finance, L9, 1-2, Room 506, 68161 Mannheim, Germany. Phone: +49-621-181-1631, E-mail: `lesnevski@bwl.uni-mannheim.de`

[2]Erasmus University Rotterdam, Erasmus School of Economics, Burgemeester Oudlaan 50, PO Box 1738, 30000 DR Rotterdam, The Netherlands. Phone: +31-10-40-81834, E-mail: `smajlbegovic@ese.eur.nl`

1 Introduction

Short selling has become an essential feature of arbitrageurs' trading strategies. The average level of shares shorted has rapidly increased over the recent years. This phenomenon has been related to an increase in capital devoted to arbitrage strategies but also to the expanded usage of short selling for market making and hedging activities. An increase in supply of lendable stocks is driven by the growth of institutional ownership (Lewellen, 2011).

This paper studies the informational role of short sellers in the equity market. A large body of recent literature has studied the ability of short sellers to detect deviations of stock prices from fundamentals and to arbitrage them away. Although most of these studies find evidence that suggests informed trading by short sellers, there are three major concerns. First, the key proxy employed to test informed trading by short sellers has been the *level* of short interest ratio (i.e., shares shorted relative to shares outstanding). However, as argued by, for instance, Asquith, Pathak and Ritter (2005), the negative relation between short interest ratio and future stock returns is also consistent with Miller's (1977) view that short-sale constraints impede trades of investors with negative beliefs, causing stocks to be overvalued. In other words, stocks that are already heavily shorted are the most difficult to short and, consequently, overvalued. Overpricing, in turn, mechanically commands future negative returns. Second, even if short interest ratio does reflect arbitrageurs' view on stock prospects, profits from following this view might be captured by high lending fees (Drechsler and Drechsler, 2016). Finally, using similar arguments as in the literature on institutional trading (e.g., Bennett, Sias and Starks, 2003), if short sellers can forecast returns, then cross-sectional variation in future returns should be related to *changes in short interest ratio* as a proxy for informed short selling. However, previous research documents, at best, inconclusive evidence on the predictive power of changes in short interest ratio for stock returns after controlling for the level of short interest ratio (Boehmer, Huszar and Jordan, 2010).

This paper proposes a new measure of *informed* short selling (or short covering). The simple intuition of the proxy is based on the distribution of the short interest ratio in the cross section of stocks. Namely, two prevalent features arise when analyzing the short interest ratio: First, the level of the short interest ratio differs dramatically across firms and it is highly persistent over time. As a consequence, informed short-selling in response to (short-term) mispricing may be hardly captured by the cross section of the short interest ratio. Second, certain companies exhibit a larger time-series variation in the share of shorted stocks than others. Therefore, even a deviation of the short interest ratio from the expected or usual level short selling may not be sufficient to extract the informed trading from the noisy signal of the short interest ratio. Namely, if two stocks exhibit the same increase in short interest ratio relative to their previous levels, we expect that an increase for a stock with a usually stable level of short interest is more informative relative to a stock with a usually high variation of short selling. These distribution differences may arise because certain stocks serve as perfect candidates for market making or hedging while other stocks attract primarily arbitrageurs (Desai, Ramesh, Thiagarajan and Balachandran, 2002; Diether, Lee and Werner, 2009). In this paper, we account for these differences across stocks and offer a novel proxy for informed short selling: the standardized unexpected short interest ratio ($SUSIR$), or simply *surprise in short interest*.

Using the surprise in short interest as a proxy for informed short selling, we find several new insights. First, we show that the information from short interest reports is not incorporated in stock prices instantly after public announcements. Stocks with top (bottom) 30% surprises in short interest experience strong price drift of around -0.25% (+0.27%) within 30 days after the dissemination. Second, we construct a monthly updated measure that captures this price drift. A portfolio strategy that buys stocks with the 10 percent lowest surprise in short interest (surprisingly high short covering) and sells stocks with the 10 percent highest surprise in short interest (surprisingly high short selling) yields a statistically significant risk-adjusted return of around 4 to 6 percent p.a. over the next month. This return spread is statistically and economically significant for both the equal- and value-weighted long-short portfolio. Most notably, the effect is present on both the long and short side of the portfolio. The predictive ability of $SUSIR$ is not captured by standard risk factors, mispricing-related anomalies, and other proxies of informed short-selling and short-selling impediments.

Third, we find that the return predictability stems from the ability of short sellers to predict changes in company's fundamental value. In particular, positive surprises in short interest predict lower unexpected earnings and lower cumulative abnormal returns around earnings announcements. Moreover, the profitability of the long-short strategy based on the surprises is particularly strong around earnings announcements when valuation-relevant news are released and these fundamental news are incorporated into prices. These findings suggest that short sellers trade on mispricing arising from biased beliefs of the overall market about companies' fundamentals.

Finally, our final analysis provides evidence that the persistence in mispricing and, as a consequence, the return predictability is partially explained by Shleifer and Vishny's (1997) limits-to-arbitrage argument. That is, general trading impediments, such as illiquidity and idiosyncratic volatility, are positively associated with the magnitude of predicted returns. Interestingly, a low supply of stocks to borrow, a common proxy for short-sale constraints (Nagel, 2005), has essentially no relation to the predictive power of the surprise in short interest, suggesting that $SUSIR$ captures the short sellers' informed trading rather than purely overpricing due to short-sale constraints.

The findings of this paper contribute to the literature in several important ways. First, by employing the surprise in short interest, we provide new evidence on informed trading by short sellers. Previous studies relate short selling to future negative stock returns (e.g., Desai et al., 2002; Cohen, Diether and Malloy, 2007; Boehmer, Jones and Zhang, 2008; Diether et al., 2009) and future changes in fundamentals (e.g., Hirshleifer, Teoh and Yu, 2011; Akbas, Boehmer, Erturk and Sorescu, 2013). Rapach, Ringgenberg and Zhou

(2016) find similar results for aggregate short interest. In particular, the authors show that detrended aggregate short interest predicts market returns and this predictability stems from the cash flow channel. Although it has been documented that the level of short selling predicts both stock returns and company fundamentals, this predictability might have different interpretations. Namely, stocks with high level of short interest tend to be more difficult to short, which mechanically results in predictability due to short-sale constraints (e.g., Asquith et al., 2005). Moreover, even if high level of short selling is an indicator of bearish views, it is associated with high short-selling risk (Engelberg, Reed and Ringgenberg, 2016) and high lending fees (Drechsler and Drechsler, 2016) casting doubt on the risk-adjusted net profitability of strategies based on the level of short interest ratio. If short sellers are informed investors and smartly enter and cover short positions, we should also expect that an increase (decrease) in short interest is also related to future negative (positive) returns. Interestingly, while changes in short interest ratio have been shown to predict future fundamentals (Deshmukh, Gamble and Howe, 2015), to the best of our knowledge, there are only weak, if any, evidence on their incremental ability to predict stock returns (Boehmer et al., 2010). In this paper, we first propose the *surprise in short interest* as a novel and intuitive proxy for informed short selling. Second, we show that this proxy predicts *both* stock returns and changes in fundamentals beyond the level of short interest and other standard determinants. Lastly, this predictability is not explained by common proxies of short-sale constraints.

We also contribute to the literature on market efficiency. In contrast to voluminous literature on earnings reports (e.g., Ball and Brown, 1968; Jones and Litzenberger, 1970; Bernard and Thomas, 1990; Mendenhall, 2004), to our knowledge, Senchack and Starks (1993) is the only paper that considers market reaction to short interest reports. The authors document negative market returns around announcements period for stocks with reported large increase in short interest. Our results extend this picture. We show that not only large increases but also large decreases in short interest, as captured by surprise in short interest, are associated with abnormal returns. More importantly, we observe a price drift for longer than 30 days after the public announcement of short interest.

In a related paper, Cohen et al. (2007) employ a detailed proprietary data set on lending activity at a daily frequency and show that shorting demand, which reflects intended changes of arbitrageurs' short positions, is a strong predictor of future stock returns. The data from that study is not publicly available and therefore could not be observed and processed by investors. In contrast, the data used in our study is published by stock exchanges on regular basis and is available to all market participants. Taking this fact into account, return predictability documented in our study questions market's ability to efficiently incorporate information from short sale reports into stock prices.

Finally, a recent strand of literature studies the existence and source of capital asset pricing anomalies (Stambaugh, Yu and Yuan, 2012; Engelberg, McLean and Pontiff, 2015; Stambaugh, Yu and Yuan, 2015; Stambaugh and Yuan, 2016; McLean and Pontiff, 2016). These anomalies may arise due to behavioral biases as a result of mispricing in combination with limits to arbitrage (Nagel, 2013; Shleifer and Vishny, 1997). Although the literature has documented that arbitrageurs, in particular short sellers, exploit well-known profitable capital market anomalies (Hanson and Sunderam, 2014; Akbas, Armstrong, Sorescu and Subrahmanyam, 2015, 2016; Jank and Smajlbegovic, 2016), we find that returns from informed short selling as proxied by the surprise in short interest cannot be entirely explained by trading on these anomalies. Therefore, we contribute to this literature stream by revealing a previously undocumented anomaly based on the informed trading by short sellers.

The remainder of the paper is structured as follows: Section 2 describes the data set, introduces the proxy of informed short selling, and provides first descriptive statistics of the variables. Section 3 presents evidence on informed trading by short sellers. In particular, the section consists of analyses on short sellers' ability to predict future stock returns and changes in companies' fundamentals. Section 4 concludes.

2 Data

2.1 Data sources

In our study, we use standard data sources. Equity market data on the stock level are obtained from CRSP, whereas accounting data come from the Compustat annual file. Information about the number of shares shorted is from the Compustat supplementary short interest file. Short interest ratio (SR) is defined as the mid-month short interest over the number of shares outstanding.[1] Other short interest-based control variables are days-to-cover ratio (Hong, Li, Ni, Scheinkman and Yan, 2016) and short interest over institutional ownership (Drechsler and Drechsler, 2016). Days-to-cover (DTC) is defined as short interest ratio over average daily turnover in the same month. To calculate short interest over institutional ownership (SR_{IO}), we divide short interest ratio by institutional ownership ratio. We retrieve institutional ownership in individual stocks from Thomson Reuters institutional (13f) holdings. Residual institutional ownership (RIO) is calculated as a residual in the cross-sectional regression of the logit-transformed institutional ownership ratio on log size and log size squared following Nagel (2005). We obtain the Corwin and Schultz (2012) spread, the bid-ask spread estimator calculated using daily high and low prices ($HLSPREAD$), directly from the authors[2]. Idiosyncratic volatility ($IVOLA$) is the standard deviation of residuals over past month in the daily regression of excess return on Fama and French (1993) three factors, as in Ang, Hodrick, Xing and Zhang (2006). Various pricing factors are obtained either from the corresponding authors' websites or calculated based on publicly available data. For instance, underpriced minus overpriced (UMO) factor is calculated following Stambaugh and Yuan (2016) using the mispricing score ($MISP$) from Robert Stambaugh's website. Analyst data including forecasts and observed earnings are collected from IBES. Market beta ($MBETA$) is the slope coefficient

[1]Starting 2007 the short interest data is reported bimonthly. For consistency, we use only mid-month figures.

[2]We thank Shane Corwin for providing these data.

in the regression of daily excess stock returns on the market factor based on the daily data over past year. Return on assets (ROA) is income before extraordinary items divided by total assets. Investment (INV) is the asset growth measure of Cooper, Gulen and Schill (2008). Other standard control variables, such as log size, log book-to-market ratio, momentum and short-term reversal, are defined as it is common in the literature. A more detailed description of variables is given in the Appendix in Table A.1.

Our sample period is from March 1980 to December 2013. The starting date of our sample is determined by the availability of reliable data on the 13F filings, an essential ingredient of one of our important control variables. To construct our universe of the U.S. equity market, we apply a number of filters. In particular, we include in our analyses stocks with share code 10 and 11. We consider all AMEX, NYSE, NASDAQ stocks. The NASDAQ sample, however, starts in June 2003, limited by availability of short interest data in Compustat. To ensure that our results are economically meaningful and not driven by penny stocks, we drop stocks with a previous month's price below 5$ and stocks below the 5th NYSE market capitalization percentile. Finally, we ensure that our sample stocks have a non-missing mispricing score, which is an important control variable in our study.

2.2 Surprise in short interest

We introduce a new measure of arbitrageurs' opinion about stock misvaluation - the surprise in short interest. This measure relies on two important ingredients. The first ingredient is based on the observation that the short interest ratio is persistent over time. There are a few potential reasons for such persistence. On the demand side, it could be driven by market making and hedging activities involving equity short selling. On the supply side, the rise of institutional investors and especially exchange traded funds boosts short selling by pushing lending fees down. Such persistent component in short interest ratio does not necessary reflect investors' views about stock future performance. We extract an unexpected component of short interest ratio by subtracting 12-months moving window mean from short interest ratio.[3] An implicit assumption is that investors rely on 12-months window to form their expectation about the level of shorting activity. The second ingredient is based on the observation that the short interest ratio is volatile and this volatility varies across stocks. Thus, an unexpected change in short interest ratio could be small but significant relative to the volatility level of short interest ratio. Or inversely, a large unexpected change in short interest ratio could be negligible if compared to volatility. To address this issue, we divide the unexpected change in short interest ratio by the volatility of short interest ratio. In particular, we use past 12-month moving window standard deviation of short interest ratio. Formally, for stock i and month t the surprise in short interest is defined as:

$$SUSIR_{i,t} = \frac{SR_t - \overline{SR}_{t-1,t-12}}{\sigma^{SR}_{t-1,t-12}}, \tag{1}$$

where $\overline{SR}_{t-1,t-12}$ is the twelve months moving window mean of short interest ratio and $\sigma^{SR}_{t-1,t-12}$ is the volatility of short interest ratio over past 12 months.[4] This proxy serves as the variable of our main interest.

2.3 Summary statistics

We start our analysis with descriptive statistics. They are presented in Table 1. Panel A reports mean, standard deviation, 1st, 10th, 50th, 90th and 99th percentiles of the variables. All variables are winsorized at 0.1% and 99.9% levels. Surprise in short interest has a mean of 0.332 and a median of 0.006 with a standard deviation of 2.069. Thus it is slightly skewed towards positive surprises. This result can be explained by the fact that short interest is limited by zero from below. Slight positive skewness is also observed for short interest ratio, days-to-cover and short interest over institutional ownership. Mean

[3] We also use different window lengths and alternative procedures, for example autoregressive models, to estimate the unexpected component of short interest ratio. The results remain qualitatively similar.

[4] This definition is somehow similar to the definition of standardized unexpected earnings surprises (SUE) that serves as a proxy for earnings surprises in accounting and finance literature (see for example Foster, Olsen and Shevlin (1984)). Such similarity in definitions explains similarity in names.

and median for these variables are 0.035 and 0.016, 6.085 and 3.721, and 0.067 and 0.033 correspondingly. Average market beta is around 1. Firm size is highly skewed. Its median value is 826 $Mio and its mean is around 4,498 $Mio. For this reason, we apply log transformation to this variable. Mispricing score ranges by construction from 0 to 100. A mispricing score close to 0 indicates stock's strong underpricing, whereas 100 indicates strong overpricing. Bid-ask spread varies from 0.2% for the 1st percentile to 2.5% for the 99th percentile. Mean institutional ownership ratio is 56.1% and increases to 100% for the top 1% stocks. Other summary results are consistent with the literature.

[Table 1 about here.]

Panel B shows Spearman correlations between variables. As expected, $SUSIR$ is related to short interest-based measures. Its correlation coefficients with SR, DTC and SR_{IO} are equal to 0.22, 0.26 and 0.26, respectively. It is of great importance to control for these variables in order to test for the incremental effect of $SUSIR$ on our dependent variables. Drechsler and Drechsler (2016) show that SR_{IO} serves as a good proxy for lending fees. Thus, $SUSIR$ tends to be only weakly correlated with shorting costs. Moreover, $SUSIR$ is not correlated with proxies for limits to arbitrage, such as size, idiosyncratic volatility, bid-ask spread and institutional ownership ratio. In contrast, the other short interest-based variables have clearly some relation. For instance, the correlation coefficients for these variables with $HLSPREAD$ and IO are 0.27, 0.13, 0.31 and 0.55, 0.22, 0.24, respectively. The results for the mispricing score go in a similar direction. Its correlation coefficient with $SUSIR$ is 0.02. In comparison, the correlation coefficients with SR, DTC and SR_{IO} are 0.13, 0.13 and 0.19, respectively. Thus, our measure is less related to the aggregate mispricing score than existing short interest-based measures.

3 Results

3.1 Surprise in short interest, short interest announcements, and stock returns

In the first step, we strive to understand how stock prices react to surprises in short interest. In particular, we take a closer look at the announcement period and investigate how fast this unexpected information on short interest is incorporated into prices.

To conduct this analysis, we rely on the announcement dates of short interest data for stocks traded primarily on NYSE from 1995 to the end of our sample period. This sample choice is defined by the availability of historical short interest dissemination dates. It is worth noting that there is a significant time gap of several days between the settlement date (when the short interest data is measured) that is always in the middle of the month and the dissemination date (when the information is published by NYSE) which usually scheduled in the last week of the corresponding month. This time gap varies over years. It is possible that some information diffuses before the official dissemination date of the exchange. Therefore, the NYSE dissemination date is the latest possible date on which the surprise in short selling is finally available to the market. Consequently, we define the period of seven trading days prior NYSE's publication as the dissemination period.

To test the price effects of the surprise in short interest, we calculate our surprise measure for each individual stock and month using the short interest data measured in the middle of the month. Then, at each NYSE dissemination date, we define one portfolio that consists of stocks with the top 30 percent surprises in short interest and one portfolio with the lowest 30 percent surprises. In the next step, at each announcement day, we calculate the abnormal and cumulative abnormal returns of the two portfolios around the announcement period using the equal-weighted market return as a benchmark. Finally, we calculate the cumulative average abnormal return across all events for each of the two portfolios.

[Insert Figure 1]

Figure 1(a) and 1(b) display the cumulative average abnormal returns of the two portfolios from 10 trading days before to 30 days after the NYSE dissemination day. Several important findings emerge. First, stocks with a high/positive surprise in short interest went through a significant price increase before the dissemination period, whereas stocks with a low/negative surprise in short interest significantly decreased before the dissemination period. Second, during the dissemination period, prices of stocks with high and low surprises start to revert relative to their previous performance. Stocks with surprising high (low) short interest start to decrease (increase) in value. Third and most interestingly, these negative *and* positive price reactions continue even after the publication of short interest data. In economic terms, stocks with high (low) values of *SUSIR*, decrease (increase) their value by 0.25% (0.27%) within 30 trading days after the NYSE dissemination day. These values are not just economically meaningful but also statistically significant at the 1% level. Thus, we observe a strong price drift, but only weak, if any, reaction to short interest report immediately on the day of the public announcement. There are a few possible explanations to this finding. First, some market participants might obtain the short sale data before the NYSE's public announcement, for instance, through commercial data providers. As a result, reaction to new information is incorporated in prices on different days within the dissemination period. Second, limits to arbitrage might prevent information to be incorporated instantly. The drift might be a result of investors avoiding stocks that are more difficult to trade or that are prone to noise trader risk (De Long, Shleifer, Summers and Waldmann, 1990; Shleifer and Vishny, 1997). However, short-sale constraints are unlikely to play a crucial role due to the fact that we even observe a *positive* price drift for stocks with low surprise in short interest. These stocks are also associated with a weak announcement effect and an investment strategy exploiting this drift does not involve any costs or risk associated with short selling. Finally, the construction of our measure requires a time series of 13 figures on short interest per stock and thus might represent significant signal extraction costs for many market participants. This explanation would be in line with the Rapach et al.'s (2016) interpretation of market return predictability using detrended aggregate short interest.

In sum, the result of positive surprises in short interest are consistent with Senchack and Starks (1993), who consider the reaction of stock prices to large increases in short interest. However, the positive price reaction after negative surprises is novel and contributes to a better understanding of the role of short selling information in financial markets. More importantly, we document a long-run price drift following announcements of surprises in short interest. In the next steps, we focus our analysis on this price drift and study the predictive power of *SUSIR* on a monthly frequency. We explore whether the return predictability can be explained by well-known risk factors or prominent anomalies. Further analyses investigate whether this return predictability reflects mispricing and delayed price response to fundamental news. Lastly, we study the role of limits to arbitrage in explaining the existence of this return predictability.

3.2 Portfolio sorts and long-short strategies

In this part, we employ the portfolio approach on a monthly frequency to test whether the surprise in short interest predicts future returns in the cross section of U.S. stocks. In particular, each month we sort the stocks into deciles according to the standardized unexpected short interest ratio which is measured in the middle and announced at the end of previous month. If *SUSIR* captures long-lived information about stock performance, we expect that stocks with a positive surprise in short interest are associated with lower abnormal returns in the next month relative to stocks with the negative surprise in short interest.

[Insert Table 2]

Table 2 reports the average excess returns of the individual deciles and risk-adjusted returns across different factor models. Panel A shows the equal-weighted portfolio returns, whereas Panel B shows the corresponding returns if stocks are weighted according to their market capitalization within the portfolios. As evident from Column (1), we find that stocks in the highest decile earn on average lower future returns compared to stocks in the lowest decile. Consistent with our expectation that *SUSIR* captures short sellers' informed trading, the difference between the two extreme equal-weighted portfolios amounts to 0.44%. This result is of meaningful economic magnitude and statistically significant at any conventional level. Moreover, the relation between the surprise in short interest and return is not only reflected in the extreme portfolios. Namely, the average return is almost monotonically decreasing in the surprise in short interest with a minor exception for the ninth and tenth portfolio. A similar return pattern emerges for the value-weighted portfolios in Panel B, though slightly more noisy in economic and statistical terms.

Next, to ensure that the above observed return difference is not explained by standard risk or mispricing-related factors, we regress the time-series of the long-short portfolio returns on a number of factors. We first employ the following risk-based factor models: The seminal Capital Asset Pricing Model (CAPM) by Sharpe (1964) and Lintner (1965), the Fama and French (1993) three-factor model, the Carhart (1997) four-factor model, the Carhart (1997) model with the Pastor and Stambaugh (2003) liquidity factor, and the two most recent models of Fama and French (2015) and Hou, Xue and Zhang (2014). The results are reported in Columns (2) to (7). Most importantly, the risk-adjusted returns are positive and statistically different from zero across all factor models, suggesting that classical risk-factors are unable to explain the predictive power of *SUSIR*.

Moreover, to ensure that the surprises in short interest do not simply mimick the arbitragers' trading on well-known mispricing factors, we also account for factors that have been recently related to misvaluation rather than traditional risk explanations. Namely, in Column (8), we employ the Carhart (1997) model including two additional factors that have been introduced as mispricing factors in the original studies: the quality-minus-junk (QMJ) (Asness, Frazzini and Pedersen, 2013) and the betting-against-beta (BAB) factor

(Frazzini and Pedersen, 2014). We choose these factors as mispricing candidates because, first, Jank and Smajlbegovic (2016) document that the two factors are associated with the actual trading of short sellers, and, second, Harvey and Liu (2016) show that those factors, especially QMJ, are of great importance for explaining the cross section of stock returns when returns are value-weighted. The alpha in Column (8) suggests that the surprise in short interest does not simply capture trading on these two prominent mispricing factors. In the last specification, we control for the recent mispricing factor proposed by Stambaugh and Yuan (2016), who combine 11 different anomalies into one score reflected in the factor underpriced-minus-overpriced (UMO). Although the adjusted return of the value-weighted long-short *SUSIR* portfolio slightly decreases relative to the raw portfolio return, it remains economically large and statistically significant.

As a robustness check, we form a more conservative long-short portfolio using the top and bottom 30 percent of stocks instead of deciles. This sorting exercise largely reduces the difference between the average surprise in short interest of the two extreme portfolios and thereby works against finding a significant return difference. Nonetheless, the average returns of the portfolio across all models are positive and statistically significant. This result demonstrates that the predictive relation between surprise in short interest and stock returns is not driven by extreme observations.

[Insert Table 3]

In the next step, we explore whether the return spread between stocks with low and high surprise in short interest can be explained by other measures related to short-sale constraints. Namely, previous empirical research finds that the level of short interest predicts future negative abnormal returns (Desai et al., 2002; Boehmer et al., 2008; Diether et al., 2009; Asquith et al., 2005). This predictive relationship has been associated with two mutually non-exclusive explanations: On the one hand, this predictability suggests informed trading by short sellers, however, on the other hand, the effect might stem also from the fact that stocks with a high level of short interest are more difficult to short, resulting in overpricing and predictability of negative future returns. To address the concern that *SUSIR* solely captures the level effect of short interest, we now include the return of the long-short portfolio based on the two extreme short interest ratio deciles to the Carhart (1997) model. The result is reported in Column (2) of Table 3. Although the short interest ratio portfolio is significantly related to our surprise portfolio, it does not entirely explain the *SUSIR* portfolio return. The alpha decreases relative to the Carhart (1997) model (Column (1)) to a still statistically significant return of around 0.255 (0.288) percent for the equal-weighted (value-weighted) portfolio.

In a recent study, Hong et al. (2016) show that days to cover (DTC), the ratio of short interest to trading volume, measures the costliness of exiting crowded trades such that arbitrageurs require a premium for trading stocks with high DTC. In Column (3) we analyze whether this premium can explain the return spread arising from the surprise in short interest. We document a positive relation of our portfolio to the DTC long-short portfolio. Also, more importantly for our study, this relation does not entirely explain the return based on surprise in short interest.

Lastly, Drechsler and Drechsler (2016) find that the ratio between short interest (demand for shorting) and shares held by institutional investors (lending supply) negatively predicts the cross section of stock returns. The authors justify this predictability as a premium that compensates short sellers for their limited risk-bearing capacity. Including a long-short portfolio based on this ratio into our time-series regression in Column (4), we find that shorting premium does not entirely explain the effect of surprise in short interest.

Finally, in the last specification, we add all three previously discussed short interest-related factors to the (Carhart, 1997) four-factor model. We observe that the effect of informed short selling, as proxied by the surprise in short interest, remains economically and statistically significant even in the full specification.

3.3 The cross section of individual stock returns

So far, the results of the portfolio sorts suggest a negative relation between surprise in short interest and future stock returns. In this part of the paper, we employ the returns of individual U.S. stocks and conduct a regression analysis along the lines of Fama and MacBeth (1973) with monthly excess returns as the dependent variable. Relative to the portfolio approach, this stock-specific approach allows to easily account for other firm characteristics and rule out other possible explanations for the effect of $SUSIR$. Formally, we run a cross-sectional regression for each month t:

$$Ret_{i,t} = \alpha_t + \beta_t SUSIR_{i,t-1} + \mathbf{x}'_{\mathbf{i,t-1}} \mathbf{b_t} + \varepsilon_{i,t}, \tag{2}$$

where $SUSIR_{i,t-1}$ denotes the stock-specific surprise in short interest as defined in Equation (1) and $\mathbf{x_i}$ represents a vector of control variables depending on the specification. All explanatory variables are standardized with a mean of zero and a standard deviation of one. Then, we calculate the time-series average of each estimated regression coefficient and its t-statistic. To account for autocorrelation and heteroskedasticity in the error terms, we use the Newey and West (1987) correction with twelve lags. If a positive surprise in short interest reflects informed short selling and has predictive power for stock returns, we expect a significant negative estimate of β. We report the estimation results for different specifications of Equation 2 in Table 4.

[Insert Table 4]

The first specification in Column 1 of Table 4 considers only $SUSIR$ and the standard control variables as explanatory variables. Using this simple design, we find that the surprise in short interest negatively predicts individual stock returns. The regression coefficient associated with the proxy is 0.114 with a corresponding t-value of 5.73. Therefore, the average return spread between a stock that is one standard deviation below the mean and a stock that is one standard deviation above the mean of *SUSIR* is around 0.23%. This result is consistent with the portfolio sort findings in Table 2 using the top and bottom 30 percent of the cross section of stocks.[5]

Next, we account for the level of short interest ratio and proxies of short-sale constraints in the regression framework. In Column (2) to Column (4), we add the level of short interest, days-to-cover, and the ratio between short interest and shares held by institutional investors, respectively. As evident from the table, for all three specifications, the coefficient of surprise in short interest slightly decreases but remains economically and statistically significant. Moreover, in line with the original studies, we find that all three control variables negatively predict the cross section of future stock returns.

Then, we test whether the two recent prominent factors, investment and profitability (Fama and French, 2015; Hou et al., 2014) affect the predictive power of *SUSIR*. In Column (5) we observe that the coefficients of both variables have the expected signs but the effect of surprise in short interest is essentially unchanged relative to our benchmark specification in Column (1).

In the next step, we account for the possibility that surprise in short interest simply reflects the trading of short sellers on well-known mispricing factors. Similar to the portfolios sorts, we rely on the mispricing score of Stambaugh and Yuan (2016). A large score indicates that a stock is overpriced, whereas a small score suggests underpricing. Therefore, we expect the coefficient to the mispricing score to be negative. As evident from Column (6), we find that the proxy of Stambaugh and Yuan (2016) negatively predicts stock returns. More important for this study, we observe that the surprise in short interest remains a significant cross-sectional predictor of future stock returns.

In Column (7) we include idiosyncratic volatility of the individual stock to control for the low-volatility anomaly. Most notably, it does not affect our result on *SUSIR*.

[5] The difference of the average $SUSIR$ between the lowest 30 percent portfolio and the highest 30 percent portfolio is very similar to the spread between the stocks one standard deviation below and above the mean $SUSIR$ value.

Though, consistent with previous research (e.g., Ang et al., 2006), we document that high idiosyncratic volatility is associated with lower future returns. Lastly, in Column (8), we include all previously introduced variables into one specification and find that the surprise in short interest is still a meaningful and important predictor of the cross section of future stock returns.

Both the portfolio approach and the analysis on the level of individual stock returns suggest that $SUSIR$ captures the informed trading by short sellers. In the remaining part of this paper, we strive to better understand the source and channels of this return predictability.

3.4 The time-series dimension of return predictability

To understand the source of return predictability, we turn to the time-series properties of the $SUSIR$ long-short strategy. We check the stability of monthly returns delivered by the strategy over time. In particular, we calculate and then analyze the cumulative returns. Severe return crashes would be consistent with a risk-based explanation of strategy's abnormal profits. Momentum (Daniel and Moskowitz, 2016) and days-to-cover (Hong et al., 2016) strategies are prominent examples of risky strategies that experienced severe crashes, in particular, around the 2008 financial crisis. Second, we look at the performance of the $SUSIR$ long-short portfolio up to 24 months after the portfolio formation. The main purpose of this exercise is to rule out the possibility that the negative relation between $SUSIR$ and future stock returns is the result of temporary price change due to short-sellers' overreaction or non-informed demand shocks. In case of these explanations, we would expect a reversal in the strategy's return in the long run.

Figure 2 illustrates the overall performance of the $SUSIR$ long-short strategy over time. We plot the logarithmic cumulative raw returns to retain the comparability of the performance across different time periods. The plot indicates that the strategy does not exhibit strong negative returns or crashes over the sample period. This finding serves as the first evidence in favor of the mispricing-based explanation of strategy's profits.[6]

[Insert Figure 2]

Is the profitability of the $SUSIR$ long-short strategy the result of temporary or permanent price changes? To answer this question we analyze the long-run performance of the $SUSIR$ buy-and-hold portfolio. First, we form decile portfolios based on $SUSIR$ in every month t. Next, we calculate equal- and value-weighted raw excess returns of the first minus tenth decile portfolios in month $t+k$, where $k \in \{1, ..., 24\}$. Finally, We run the following time-series regression for each holding period month k:

$$\begin{aligned} LS\_SUSIR_{t+k} = \alpha_k + \beta_{MKTRF,k} MKTRF_{t+k} + \beta_{SMB,k} SMB_{t+k} + \\ + \beta_{HML,k} HML_{t+k} + \beta_{UMD,k} UMD_{t+k} + \varepsilon_{k,t+k}, \end{aligned} \tag{3}$$

where LS_SUSIR_{t+k} is the raw excess return in month $t+k$ of the long-short portfolio formed in month t and $MKTRF_{t+k}$, SMB_{t+k}, HML_{t+k}, UMD_{t+k} are returns on the four factors of Carhart (1997) model in month $t+k$. The intercept of the regression (α_k) is the alpha of the buy-and-hold strategy k months after portfolio formation.

[Insert Figure 3]

Figure 3 presents equal-weighted and value-weighted average holding period alpha obtained by accumulating over holding period months the four-factor alphas from Equation 3. The figure reveals a pattern that is consistent with $SUSIR$ predicting permanent price changes. The long-short portfolio delivers abnormal returns up to 18 months after portfolio formation and the returns do not revert, as it would be the case for temporary price changes. Surprisingly, the performance of equal-weighted and value-weighted portfolios are very similar until month 8. After this point, the cumulative alpha of value-weighted portfolio

[6] In unreported results we examine the impact of Baker and Wurgler's (2006) investor sentiment index on the returns of $SUSIR$-based long-short strategy. Our results reveal no significant relation.

fluctuates around 2%, whereas equal-weighted portfolio continues to deliver abnormal performance until month 18 when it reaches 3.6 %. Thus, the ability of surprise in short interest to predict stock returns is not limited to first months and is not associated with price reversal in the long run.

3.5 Surprise in short interest, biased expectations and fundamental news

In this section we ask whether the ability of surprise in short interest to predict the cross section of stock returns is the manifestation of risk or mispricing. Return predictability might arise due to biased expectations of market participants. However, when the information becomes available to the market, the firm's value should converge to its fundamental level. To be consistent with this type of mispricing-based explanation, $SUSIR$ should satisfy the following two conditions. First, it should predict future changes in fundamentals (cash-flow news). Second, the ability to predict stock returns should be stronger on the days when such valuation-relevant news are released. We use quarterly earnings announcements to test our two hypotheses. Companies' quarterly earnings reports contain valuable information about firm fundamentals. We use earnings surprises as a proxy for fundamental news in these reports. Market participants update their expectations upon report releases. Consequently, these changes in expectations are reflected in equity prices.

To test our first hypothesis we ask whether surprises in short interest can predict future surprises in announced earnings. Using data on 196,719 earnings announcements, we run a panel regression of earnings surprises on the most recent $SUSIR$ measure and other controls including month-fixed effects:

$$Earnings\_Surprise_{i,t} = \alpha_t + \beta_t SUSIR_{i,t-1} + \mathbf{x}'_{\mathbf{i,t-1}}\mathbf{b_t} + \varepsilon_{i,t}, \tag{4}$$

where $SUSIR_{i,t-1}$ is the stock-specific surprise in short interest in month $t-1$ and $\mathbf{x_{i,t-1}}$ is a vector of control variables depending on the specification.

We use three measures of earnings surprises. The first measure is standardized unanticipated earnings based on past earnings (SUE^{PE}). This measure is equal to the difference between current earnings per share and earnings per share reported 4 quarter ago divided by the standard deviation of this difference over past 8 quarters. It is derived under the assumption that earnings follow seasonal a random walk and it performs well in capturing earnings news (Foster et al., 1984). The second measure is standardized unanticipated earnings based on analyst forecasts (SUE^{AF}), also known as analyst forecast errors. In contrast to SUE^{PE}, this measure uses analyst forecasts as a proxy for the market expectation of EPS. It is defined as actual EPS net of the most recent mean analyst forecasts over the standard deviation of analyst forecasts. This measure is obtained directly from IBES. Finally, the third measure, CAR, is the cumulative market-adjusted return over the earnings announcement window $[-1, 1]$. CRSP value-weighted portfolio is used as a benchmark index. In some specifications we control for two measures of stock misvaluation, short interest ratio and mispricing score. There are two reasons to include short interest ratio. First, short interest ratio predicts stock returns and its predictive power comes also from the fundamental news channel (Akbas et al., 2013). Second, $SUSIR$ is derived using information from short interest ratio. We also control for mispricing score that succeeds in aggregating misvaluation-related information (Stambaugh et al., 2015; Stambaugh and Yuan, 2016). If surprise in short interest is not simply loading on known factors, it should have significant predictive power on top of these factors. The explanatory variables are standardized to have zero mean and unit standard deviation for the comparability of coefficients.

Regression results are reported in Table 5. The sample period starts in May 1985 for SUE^{PE}, in May 1992 for SUE^{AF}, and in May 1980 for CAR. The sample period is limited by data availability. All earnings surprises variables are multiplied by 100 to improve table readability. All earnings surprises measures are winsorized at 1% and 99% levels. The standard errors are double-clustered by month and stock. The standard control variables are market beta, the log size, log book-to-market ratio, momentum and short-term reversal. The left panel of the table reports results for SUE^{PE}. The coefficient on $SUSIR$ in the specification that includes only month-fixed effects is -0.0302 with a t-statistic of -3.51. The negative sign of the coefficient means that an increase in surprise in short interest predicts, all else being equal, lower earnings surprises. The inclusion of standard controls makes the predictive power of $SUSIR$ even stronger (Column (2)). Controlling for $MISP$ and SR slightly decreases the significance of $SUSIR$ (Columns (3) and (4)). As expected, SR and $MISP$ are also significant predictors of earnings surprises. Coefficients on other control variables are in line with the literature. Regression results for SUE^{AF} are presented in the middle panel. Two additional controls are added: number of analysts producing the forecast ($NUMEST$) and standard deviation of forecasts ($STDEV$). The coefficients on the variables are qualitatively similar to those for SUE^{PE}. Regression with CAR as the dependent variable produces statistically weaker but nevertheless significant results. They are reported in the right panel. Regression coefficient allows easy interpretation of the economic importance of $SUSIR$. In the specification with solely month-fixed effects, the coefficient -0.0515 (t-stat -3.08) means that a two standard deviations increase in $SUSIR$ results in, all else being equal, 10.3 basis points lower announcement returns. Accounting for $MISP$ and SR slightly decreases this number to 7.3 basis points. The statistical significance of $SUSIR$ stays above the 95% confidence level in all considered specifications. Interestingly, the predictive power of SR is much stronger when using CAR as a measure of earnings surprises. It has more than double the predictive power of $SUSIR$ and even stronger than for $MISP$. Overall, the tests show that surprise in short interest is a statistically and economically significant predictor of earnings surprises. This result is in line with our first hypothesis.

[Insert Table 5]

To test the second hypothesis we turn to a daily frequency and adopt the methodology of Engelberg et al. (2015). Sample contains 12,552,943 day-stock observations. Earnings announcement dates are from the Compustat quarterly file. Our goal is to compare the predictive power of short interest surprise on announcement days to non-announcement days. We first define an earnings announcement period dummy (EAP) equal to one for the days that are within the one day window around earnings announcement. We run a panel regression of raw daily stock returns on EAP, the most recent $SUSIR$, their interaction ($SUSIR \times EAP$) and other control variables:

$$Ret_{i,t} = \alpha_t + \beta_{1,t} EAP_{i,t} + \beta_{2,t} SUSIR_{i,t-1} + \beta_{3,t} SUSIR_{i,t-1} \times EAP_{i,t} + \mathbf{x}'_{\mathbf{i,t-1}} \mathbf{b_t} + \varepsilon_{i,t}, \quad (5)$$

The coefficient on $SUSIR_{i,t-1}$ reflects the average predictive power of $SUSIR$ on non-announcement days. The coefficient on the interaction variable shows an additional predictive power of $SUSIR$ upon earnings announcements. The mispricing-based explanation commands a significant negative coefficient on this variable.

We report estimation results in Table 6. In all specifications, standard errors are clustered by month. Regression specification in Column (1) includes only EAP, $SUSIR$ and their interaction as explanatory variables together with day-fixed effects. Significantly positive coefficient on EAP is the manifestation of earnings announcement premium firstly discovered by Beaver (1968). $SUSIR$ is a significant predictor of stock returns on both announcement and non-announcement days. However, the predictability is almost five times stronger upon announcements. The introduction of four lags of return, return squared and daily turnover has no significant impact (Column (2)). Controlling for mispricing and short interest ratio slightly weakens but does not change the implications (Column (3)). Consistently with the mispricing-related nature of short interest ratio (Akbas et al., 2013) and mispricing score (Stambaugh et al., 2015), the interaction coefficients of these variables with EAP are also significant and imply around 3.5 times stronger return predictability upon news arrival. Further specifications address some possible risk-based explanations of the effect. Savor and Wilson (2016) argue that announcing firms have higher exposure to fundamental risk than non-announcing firms. To control for this risk heterogeneity, we include day-EAP fixed effects. As evident from Column (4), the announcement risk premium does not drive our results. In Column (5), we analyze whether the increase in exposure to systematic risk for earnings announcers documented by Patton and Verardo (2012) might influence our inference. That is, we test whether stocks with more negative surprises in short interest (associated with higher future returns) experience a larger increase in market beta around earnings announcements. For that, we introduce market return, proxied by CRSP value-weighted index (MKT), to our baseline specification from Column (2). We interact MKT with EAP and $SUSIR$ and add triple interaction of these variables. The coefficient on MKT confirms that average market beta is around 1. The coefficient on $MKT \times EAP$ reveals the average effect of earnings announcements on market beta to be positive but statistically insignificant (t-stat of 0.54). The insignificant coefficient on $MKT \times SUSIR$ of -0.00567 (t-stat of -1.27) suggests that market exposure is not significantly correlated with $SUSIR$. Surprisingly, the coefficient on triple interaction is positive and significant, meaning that high (low) $SUSIR$ stocks experience increase (decrease) in market beta on announcement days. The magnitude is economically significant. Given that $SUSIR$ is normalized to have zero mean and unit standard deviation, the coefficient of 0.0204 means that a stock whose $SUSIR$ is one standard deviation above (below) its average experiences increase (decrease) in its market beta by 0.02 around earnings announcements. $SUSIR$ long-short strategy is long in low-$SUSIR$ stocks and short in high-$SUSIR$ stocks, meaning that the market exposure of this strategy actually decreases on event days. Thus, if anything, the strategy gets less risky on these days. All in all, our tests shows that $SUSIR$'s predictive ability is economically and statistically stronger on earnings announcement days. This effect is not explained by earnings announcement premium or an increase in exposure to market risk.

[Insert Table 6]

To conclude, in this subsection we find that information contained in standardized unexpected short interest ratio is relevant for identifying misvalued stocks. We formulate two hypotheses consistent with the informational advantage of arbitrageurs. The first hypothesis states that the return predictability should be the result of biased expectations, which are corrected upon news arrival. The second hypothesis states that if biased expectations are the main channel of return predictability, then return predictability should be stronger on days when new valuation-relevant information is released. We find support for both hypotheses. First, we show that stocks with higher (lower) surprises in short interest experience lower (higher) earnings surprises. This conclusion holds for various measures of earnings surprises. Second, we find that the ability of $SUSIR$ to predict stock returns becomes almost five-fold stronger around earnings announcements. Moreover, this effect is not explained by an increase in systematic risk. Thus, our results are consistent with the mispricing-based explanation of $SUSIR$'s ability to predict stock returns.

3.6 Limits to arbitrage

Our results so far indicate that $SUSIR$ reflects information on misvaluation. Consequently, the following natural question arises: Why do investors not arbitrage away this misvaluation? Certain limits to arbitrage might sustain short-term deviations of stock prices from fundamental values. The seminal work of Shleifer and Vishny (1997) provides an important framework and justification for the persistence of mispricing and predictability of stock returns. The authors argue that arbitrage opportunities should vanish immediately as a high number of investors take positions against the mispricing, driving the stock price to its fundamental value. However, in reality, due to noise trading, the stock price might diverge in short-run even further from the fundamental value inducing losses to the arbitrageur. This fact could prevent the investors from trading on mispricing in the first place and set certain limits to arbitrage.

To test whether the predictability, at least, partially arises because of these limits to arbitrage Shleifer and Vishny (1997), we define variables that have been related in previous research to trading impediments. Then, for each of the variable, we sort the stocks into quintiles. Finally, we include dummy variables for each quintile, except for the first one, and their interactions with $SUSIR$ into a Fama and MacBeth (1973) regression framework:

$$\begin{aligned} Ret_{i,t} = \alpha_t + \beta_1 SUSIR_{i,t-1} + \sum_{k=2}^{5} \beta_k M_{Quintile=k,i,t-1} + \\ + \sum_{k=2}^{5} \gamma_k SUSIR_{i,t-1} \times M_{Quintile=k,i,t-1} + \mathbf{x_i'b} + \varepsilon_{i,t}, \end{aligned} \tag{6}$$

where $M_{Quintile=k}$ denotes the dummy variable equal to one if the limits-to-arbitrage variable M is in the kth quintile. The coefficients γ_2 to γ_5 are of main interest in this subsection. In particular, we expect that the negative predictability associated with $SUSIR$ is the strongest in the quintile with the highest limits to arbitrage. Also, note that the estimate β_1 denotes the effect of $SUSIR$ in the lowest quintile of variable M. In the following, we consider three different variables that are closely related to the mechanism of limits to arbitrage and commonly used in the literature: The spread estimator of Corwin and Schultz (2012) as a proxy for illiquidity, idiosyncratic volatility as a proxy for arbitrage risk (Pontiff, 2006), and residual institutional ownership as a proxy for short-sale constraints (Nagel, 2005). If limits to arbitrage are important for the persistence of mispricing, we expect that the predictability is the strongest (more negative) for stocks with high illiquidity, high idiosyncratic volatility, and lower residual institutional ownership.

[Insert Table 7]

Column (1) of Table 7 shows the time-series average of the cross-sectional regression coefficients employing the standard control variables, surprise in short interest, dummy variable for each quintile of illiquidity (except the first one), and the interaction terms of

$SUSIR$ with the dummy variables. In line with our prediction, we find that the predictive power of surprise in short interest is the strongest for the highest quintile of illiquidity. In economic magnitudes, the return spread associated with surprise in short interest is around three times larger for the most illiquid stocks relative to the most liquid stocks. Moreover, except for the second quintile, this effect of illiquidity is monotonic across the quintiles. The intuition behind this finding is the following: the higher the illiquidity of a stock, the slower and the more costly it is traded on the market. These additional costs could prevent investors from fully exploiting arbitrage opportunities and taking advantage of the return predictability.

Next, we examine the role of idiosyncratic volatility for the predictability of returns. In the framework of Shleifer and Vishny (1997), stocks with higher volatility are less attractive to arbitrageurs and exhibit larger predictable returns than stocks with lower volatility. Column (2) shows results in line with this hypothesis. The predictability in the quintile with the most volatile stocks is five times higher relative to the least volatile stocks. Interestingly, this relation is monotonic and the predictability using $SUSIR$ is statistically significant across four out of five volatility quintiles.

Lastly, we employ the residual of institutional ownership as a limits-to-arbitrage proxy. More specifically, Nagel (2005) suggests that this variable serves as a meaningful proxy for the supply of stocks to borrow in the lending market. In case of low supply, arbitrageurs face higher impediments to short the stock, resulting in stronger predictability in case of overpricing. In Column (3), we do not find any evidence for stronger return predictability for stocks with low lending supply. The effect is essentially the same across all five rio quintiles. There are two mutually non-exclusive possible explanations for this non-finding. First, the supply of stocks to borrow is particularly important for anomalies that are driven by the short leg. Given that the strategy based on $SUSIR$ yields profitable returns on both the long and short leg, short-sale constraints cannot represent the ultimate explanation in the first place. Second, the construction of the surprise in short interest is based on abnormal changes of short interest relative to the usual variation in the past. Consequently, an extreme realized negative or positive change in the level of short interest implies that this particular stock is most probably *not* short-sale constrained.[7]

Overall, in line with Shleifer and Vishny's (1997) limits-to-arbitrage argument, we find evidence that general trading impediments (Gromb and Vayanos, 2010), such as illiquidity and idiosyncratic volatility, are positively related to the strength of predictability. However, we also document that short-sale constraints cannot explain the effect of $SUSIR$.

4 Conclusion

This paper contributes to the ongoing discussion about the impact of short sellers on the informational efficiency of capital markets. We introduce a new measure of informed trading, surprise in short interest, that incorporates two regularities in the short interest data: strong persistence in the levels of short interest ratio and large cross-sectional differences in volatility. The long-short strategy based on this measure delivers up to 6% in annualized risk-adjusted returns that are not explained by standard stock characteristics, relation to other known short interest-based strategies or short-sale constraints. Evidence suggests that our measure identifies market mispricing that stems from biased beliefs of market participants and persists due to trading impediments, such as illiquidity and idiosyncratic volatility. Thus, surprise in short interest represents a mispricing-related anomaly that is not documented in the prior literature.

Our findings have wide-reaching implications for future studies on the informational role of short sellers. For instance, one important question remains: What is the source of arbitrageurs' informational advantage? Therefore, an interesting venue for future research is to study the relation of the surprise in short interest to future corporate events and

[7]In untabulated tests we find similar results for other short-sale constraint proxies such as the level of short interest ratio (Asquith et al., 2005) or the short interest relative to institutional shares (Drechsler and Drechsler, 2016)

news (e.g., Engelberg, Reed and Ringgenberg, 2012), insider trades, and capital market anomalies.

Appendix:

Table A.1:
Definitions of Variables

Variable:	Description:	Source:
SR	Short interest ratio is the mid-month reported short interest divided by shares outstanding. To aggregate on PERMNO level, we sum short interest over global issue key. We use the version of short interest variable from Compustat supplementary file that is not adjusted for stock splits. Short interest for NASDAQ stocks is available starting June 2003.	CRSP/Compustat
$SUSIR$	Standardized unanticipated short interest ratio defined as $\frac{SR_t - \overline{SR}_{t-1,t-12}}{\sigma^{SR}_{t-1,t-12}}$, where $\overline{SR}_{t-1,t-12}$ is 12-months moving window mean and $\sigma^{SR}_{t-1,t-12}$ is 12-months moving window standard deviation of short interest ratio. $SUSIR$ is set to missing if less than 5 observations of short interest ratio are available.	CRSP/Compustat
DTC	Days-to-cover equals to short interest divided by daily turnover.	CRSP/Compustat
SR_{IO}	Short interest over institutional ownership equals to short interest ratio over institutional ownership ratio.	Compustat/13F
$MBETA$	Market beta is a slope coefficient in the time series regression of the stock's return on the market excess return (MKTRF), with a rolling window of 252 trading days.	CRSP
LN_SIZE	Log market capitalization is calculated as the number of shares outstanding times price per share (in $Mio).	CRSP
LN_BM	Log book-to-market ratio is calculated following Davis, Fama and French (2000). The book-to-market ratio in year t is the total book value at the end of fiscal year ending in year $t-1$ divided by total market capitalization on the last trading day of the calendar year $t-1$, as reported by CRSP. The total book value is the book value of stockholders' equity, plus balance sheet deferred taxes and investment tax credit minus the book value of preferred stock. To estimate the book value of preferred stock, we use the redemption, liquidation , or par value, in this order (depending on data availability).	CRSP/Compustat
RET_MOM	Return momentum is the cumulative return from month $t-12$ to $t-2$.	CRSP

Continued on next page

Table A.1 – *Continued from previous page*

Variable:	**Description:**	**Source:**
RET_REV	Return reversal is the return over the month $t-1$.	CRSP
ROA	Return on assets equals to to income before extraordinary items over assets.	Compustat
INV	Assets growth is defined as $\frac{TA_{t-1}-TA_{t-2}}{TA_{t-2}}$ following Cooper et al. (2008).	Compustat
$IVOLA$	Idiosyncratic volatility is defined as the standard deviation of the recent month's daily residuals obtained in the regression of the excess stock returns on Fama-French 3-factors, with a rolling window of 252 trading days.	Compustat
$HLSPREAD$	Bid-ask spread of Corwin and Schultz (2012)).	Authors
RIO	Residual institutional ownership is the residual from the following monthly cross-sectional regression: $ln(\frac{IO_{i,t}}{1-IO_{i,t}}) = \alpha_t + \beta_{1,t} \times ln(SIZE_{i,t}) + \beta_{2,t} \times ln(SIZE_{i,t})^2 + \epsilon_{i,t}$. Being reported once per quarter, insitutional ownership is assumed to be constant over three-month period prior to the next report.	CRSP/13F
SUE^{PE}	Standardized unanticipated earnings surprises are defined as a forecast error in quarter t divided by the volatility of forecast error over last 8 quarter. Forecast error is calculated as the difference between EPS announced at t and EPS four quarters ago. Thus, $SUE^{PE} = \frac{EPS_t - EPS_{t-4}}{\sigma^{\Delta EPS}_{t-1,t-8}}$.	IBES
SUE^{AF}	Earnings forecasts error is equal to EPS announced at month t net of the most recent mean analyst forecast divided by the standard deviation of the most recent analyst forecasts.	IBES
CAR	Cumulative abnormal return over 1-day earnings announcement window is defined as $\sum_{t=-1}^{1}(ret_t - MKT_t)$, where MKT_t is the CRSP value-weighted portfolio.	CRSP/Compustat
$NUMEST$	Number of analyst forecasts used to calculate mean forecast.	IBES
$STDEV$	Dispersion in analyst earnings estimates.	IBES
EAP	Earnings announcement period dummy is equal to one for a one-day window around earnings announcement.	Compustat

References

Akbas, F., Armstrong, W. J., Sorescu, S. and Subrahmanyam, A. (2015), 'Smart money, dumb money, and capital market anomalies', *Journal of Financial Economics* **118**(2), 355–382.

Akbas, F., Armstrong, W. J., Sorescu, S. and Subrahmanyam, A. (2016), 'Capital Market Efficiency and Arbitrage Efficacy', *Journal of Financial and Quantitative Analysis* **51**(02), 387–413.

Akbas, F., Boehmer, E., Erturk, B. and Sorescu, S. M. (2013), Short Interest, Returns, and Fundamentals, SSRN Working Paper.

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

Asness, C. S., Frazzini, A. and Pedersen, L. H. (2013), Quality Minus Junk, SSRN Working Paper.

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

Baker, M. and Wurgler, J. (2006), 'Investor Sentiment and the Cross-Section of Stock Returns', *The Journal of Finance* **61**(4), 1645–1680.

Ball, R. and Brown, P. (1968), 'An Empirical Evaluation of Accounting Income Numbers', *Journal of Accounting Research* **6**(2), 159–178.

Beaver, W. H. (1968), 'The Information Content of Annual Earnings Announcements', *Journal of Accounting Research* **6**, 67–92.

Bennett, J. A., Sias, R. W. and Starks, L. T. (2003), 'Greener Pastures and the Impact of Dynamic Institutional Preferences', *Review of Financial Studies* **16**(4), 1203–1238.

Bernard, V. L. and Thomas, J. K. (1990), 'Evidence that stock prices do not fully reflect the implications of current earnings for future earnings', *Journal of Accounting and Economics* **13**(4), 305–340.

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

Boehmer, E., Jones, C. M. and Zhang, X. (2008), 'Which Shorts Are Informed?', *The Journal of Finance* **63**(2), 491–527.

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

Cohen, L., Diether, K. B. and Malloy, C. J. (2007), 'Supply and Demand Shifts in the Shorting Market', *The Journal of Finance* **62**(5), 2061–2096.

Cooper, M. J., Gulen, H. and Schill, M. J. (2008), 'Asset Growth and the Cross-Section of Stock Returns', *The Journal of Finance* **63**(4), 1609–1651.

Corwin, S. A. and Schultz, P. (2012), 'A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices', *The Journal of Finance* **67**(2), 719–760.

Daniel, K. and Moskowitz, T. J. (2016), 'Momentum crashes', *Journal of Financial Economics* **122**(2), 221–247.

Davis, J. L., Fama, E. F. and French, K. R. (2000), 'Characteristics, Covariances, and Average Returns: 1929 to 1997', *The Journal of Finance* **55**(1), 389–406.

De Long, J. B., Shleifer, A., Summers, L. H. and Waldmann, R. J. (1990), 'Noise Trader Risk in Financial Markets', *Journal of Political Economy* **98**(4), 703–738.

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 Finance* **57**(5), 2263–2287.

Deshmukh, S., Gamble, K. J. and Howe, K. M. (2015), 'Short selling and firm operating performance', *Financial Management* **44**(1), 217–236.

Diether, K. B., Lee, K.-H. and Werner, I. M. (2009), 'Short-Sale Strategies and Return Predictability', *Review of Financial Studies* **22**(2), 575–607.

Drechsler, I. and Drechsler, Q. F. S. (2016), The Shorting Premium and Asset Pricing Anomalies, SSRN Working Paper.

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), 260–278.

Engelberg, J., McLean, R. D. and Pontiff, J. (2015), Anomalies and News, SSRN Working Paper.

Engelberg, J., Reed, A. V. and Ringgenberg, M. (2016), Short selling risk, SSRN Working Paper.

Fama, E. F. and French, K. R. (1992), 'The Cross-Section of Expected Stock Returns', *The Journal of Finance* **47**(2), 427–465.

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

Fama, E. F. and French, K. R. (2015), 'A five-factor asset pricing model', *Journal of Financial Economics* **116**(1), 1–22.

Fama, E. F. and MacBeth, J. D. (1973), 'Risk, Return, and Equilibrium: Empirical Tests', *Journal of Political Economy* **81**(3), 607–636.

Foster, G., Olsen, C. and Shevlin, T. (1984), 'Earnings Releases, Anomalies, and the Behavior of Security Returns', *The Accounting Review* **59**(4), 574–603.

Frazzini, A. and Pedersen, L. H. (2014), 'Betting against beta', *Journal of Financial Economics* **111**(1), 1–25.

Gromb, D. and Vayanos, D. (2010), 'Limits of Arbitrage', *Annual Review of Financial Economics* **2**(1), 251–275.

Hanson, S. G. and Sunderam, A. (2014), 'The Growth and Limits of Arbitrage: Evidence from Short Interest', *Review of Financial Studies* **27**(4), 1238–1286.

Harvey, C. R. and Liu, Y. (2016), Lucky Factors, SSRN Working Paper.

Hirshleifer, D., Teoh, S. H. and Yu, J. J. (2011), 'Short Arbitrage, Return Asymmetry, and the Accrual Anomaly', *Review of Financial Studies* **24**(7), 2429–2461.

Hong, H. G., Li, F. W., Ni, S. X., Scheinkman, J. A. and Yan, P. (2016), Days to Cover and Stock Returns, SSRN Working Paper.

Hou, K., Xue, C. and Zhang, L. (2014), 'Digesting Anomalies: An Investment Approach', *Review of Financial Studies* p. hhu068.

Jank, S. and Smajlbegovic, E. (2016), Dissecting Short-Sale Performance: Evidence from Large Position Disclosures, SSRN Working Paper.

Jones, C. P. and Litzenberger, R. H. (1970), 'Quarterly Earnings Reports and Intermediate Stock Price Trends', *The Journal of Finance* **25**(1), 143–148.

Lewellen, J. (2011), 'Institutional investors and the limits of arbitrage', *Journal of Financial Economics* **102**(1), 62–80.

Lintner, J. (1965), 'The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets', *The Review of Economics and Statistics* **47**(1), 13–37.

McLean, R. D. and Pontiff, J. (2016), 'Does Academic Research Destroy Stock Return Predictability?', *The Journal of Finance* **71**(1), 5–32.

Mendenhall, R. R. (2004), 'Arbitrage Risk and Post-Earnings-Announcement Drift', *The Journal of Business* **77**(4), 875–894.

Miller, E. M. (1977), 'Risk, Uncertainty, and Divergence of Opinion', *The Journal of Finance* **32**(4), 1151–1168.

Nagel, S. (2005), 'Short sales, institutional investors and the cross-section of stock returns', *Journal of Financial Economics* **78**(2), 277–309.

Nagel, S. (2013), 'Empirical Cross-Sectional Asset Pricing', *Annual Review of Financial Economics* **5**(1), 167–199.

Newey, W. K. and West, K. D. (1987), 'Hypothesis Testing with Efficient Method of Moments Estimation', *International Economic Review* **28**(3), 777–787.

Pastor, L. and Stambaugh, R. F. (2003), 'Liquidity Risk and Expected Stock Returns', *Journal of Political Economy* **111**(3), 642–685.

Patton, A. J. and Verardo, M. (2012), 'Does Beta Move with News? Firm-Specific Information Flows and Learning about Profitability', *Review of Financial Studies* **25**(9), 2789–2839.

Pontiff, J. (2006), 'Costly arbitrage and the myth of idiosyncratic risk', *Journal of Accounting and Economics* **42**(1–2), 35–52.

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

Savor, P. and Wilson, M. (2016), 'Earnings Announcements and Systematic Risk', *The Journal of Finance* **71**(1), 83–138.

Senchack, A. J. and Starks, L. T. (1993), 'Short-Sale Restrictions and Market Reaction to Short-Interest Announcements', *The Journal of Financial and Quantitative Analysis* **28**(2), 177.

Sharpe, W. F. (1964), 'Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk', *The Journal of Finance* **19**(3), 425–442.

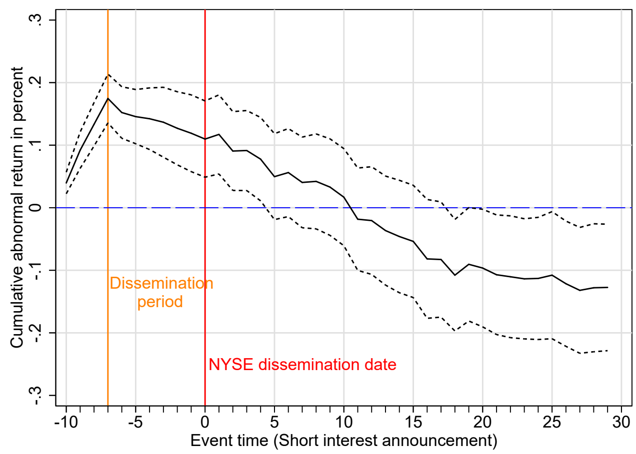
Shleifer, A. and Vishny, R. W. (1997), 'The Limits of Arbitrage', *The Journal of Finance* **52**(1), 35–55.

Stambaugh, R. F., Yu, J. and Yuan, Y. (2012), 'The short of it: Investor sentiment and anomalies', *Journal of Financial Economics* **104**(2), 288–302.

Stambaugh, R. F., Yu, J. and Yuan, Y. (2015), 'Arbitrage Asymmetry and the Idiosyncratic Volatility Puzzle', *The Journal of Finance* **70**(5), 1903–1948.

Stambaugh, R. F. and Yuan, Y. (2016), Mispricing Factors, SSRN Working Paper.

(a) CAAR for stocks with high surprise in short interest



(b) CAAR for stocks with low surprise in short interest

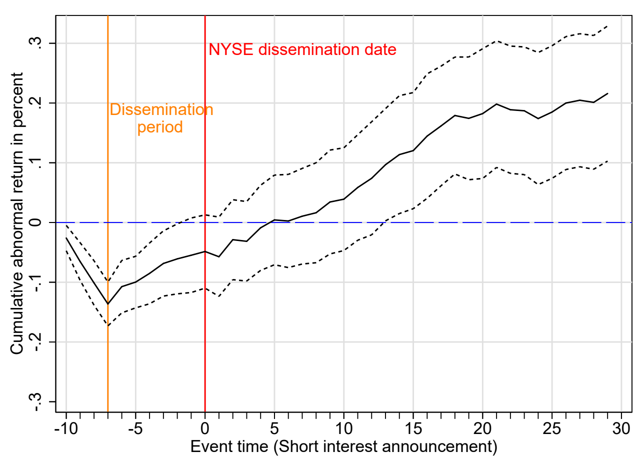


Figure 1:
Cumulative Average Abnormal Returns Around Short Interest Announcement
This figure displays the cumulative average abnormal returns (CAARs) of two portfolios based on the surprise in short interest. The benchmark is the market return and the time window is 10 days before and 30 days after the NYSE's short interest dissemination day. Figure 1(a) plots the CAAR for stocks with the highest 30% of surprises in short interest at each event day, whereas Figure 1(b) plots the CAAR for stocks with the lowest 30% of surprises in short interest at each event day. The dashed lines represent the upper and lower 90% confidence intervals. The sample starts in January 1995 and contains stocks traded on the NYSE.

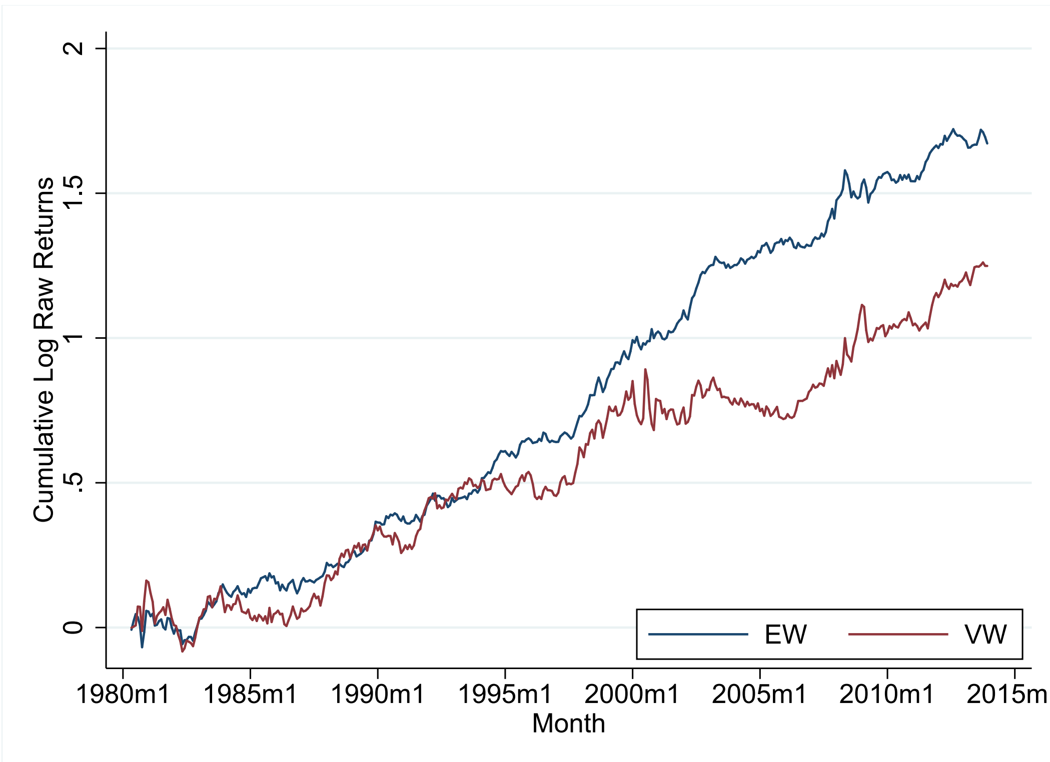


Figure 2:
Cumulative Logarithmic Raw Returns
Plotted are the monthly cumulative sum of log raw returns for equal-weighted and value-weighted $SUSIR$ long-short strategy over the period from March 1980 to December 2013.

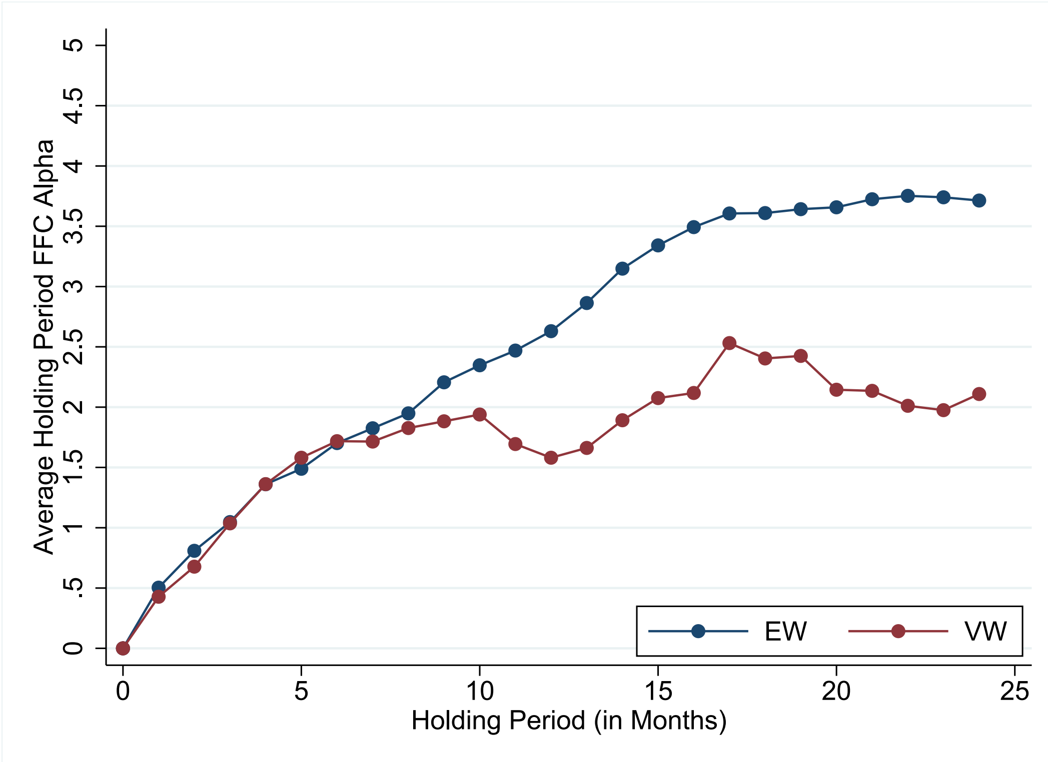


Figure 3:
Holding Period Performance of $SUSIR$ Long-Short Strategy
This figure plots the average cumulative Carhart (1997) four-factor alpha of long-short portfolio over the holding period. We first construct each month the long-short portfolio according to $SUSIR$ and calculate monthly returns in the month $t+k$, where $k \in \{1, ..., 24\}$. Second, we run a time-series regression for each holding period month k of $SUSIR$ long-short strategy on four factors. The corresponding strategy average four-factor alpha at month k is the intercept (α_k) in the following regression:

$$LS\_SUSIR_{t+k} = \alpha_k + \beta_{MKTRF,k} MKTRF_{t+k} + \beta_{SMB,k} SMB_{t+k} + \\ + \beta_{HML,k} HML_{t+k} + \beta_{UMD,k} UMD_{t+k} + \epsilon_{k,t+k}.$$

In the final step, the alphas are accumulated over holding period months. The sample period is from March 1980 to December 2013.

Table 1:
Descriptive Statistics

Panel A of this table reports univariate summary statistics (mean, standard deviation and 1st, 10th, 50th, 90th, 99th percentiles) of the variables used in this study. The first set of variables in Panel A is based on short interest data. Standardized unexpected short interest ratio ($SUSIR$) is the short interest ratio net of its 12-months moving average over its 12-months moving standard deviation. Short interest ratio (SR) is the short interest over shares outstanding. Days-to-cover measure (DTC) is equal to short interest divided by daily turnover. SR_{IO} is short interest ratio over institutional ownership ratio. The second set of variables in Panel A is the set of stock characteristics that are known to predict stock returns. Market beta ($MBETA$) is calculated using 12 months of daily data. Size ($SIZE$) and book-to-market ratio (BM) are constructed as in Fama and French (1992). Return reversal (RET_RV) is the return over the month $t-1$. Return momentum (RET_MOM) is the cumulative return from month $t-12$ to $t-2$. Investment (INV) is the asset growth measure of Cooper et al. (2008). Return on assets (ROA) is equal to income before extraordinary items over assets. $MISP$ is the mispricing score of Stambaugh et al. (2015). The third set of variables in Panel A is the set of proxies for the limits to arbitrage. Idiosyncratic volatility ($IVOLA$) is constructed using 12 months of daily data as in Ang et al. (2006). $HLSPREAD$ is the bid-ask spread of Corwin and Schultz (2012). Institutional ownership ratio (IO) is the share of shares outstanding owned by institutional investors. All variables are winsorized at 0.1% and 0.99% levels. Panel B of this table reports Spearman correlations of variables.

Panel A: Summary Statistics

			Percentiles				
Variable	Mean	SD	1st	10th	Median	90th	99th
$SUSIR$	0.332	2.069	-2.935	-1.484	0.006	2.338	6.452
SR	0.035	0.049	0.000	0.001	0.016	0.091	0.239
DTC	6.085	7.509	0.015	0.496	3.721	13.951	37.831
SR_{IO}	0.067	0.122	0.000	0.003	0.033	0.153	0.498
$MBETA$	1.042	0.452	0.018	0.505	1.018	1.607	2.283
$SIZE$	4498.113	17182.012	33.312	126.298	826.032	8202.445	69739.656
BM	0.643	0.528	0.000	0.190	0.545	1.169	2.338
RET_RV	0.012	0.113	-0.291	-0.113	0.009	0.138	0.342
RET_MOM	0.196	0.518	-0.606	-0.280	0.121	0.688	2.126
INV	0.158	0.397	-0.327	-0.065	0.081	0.399	1.722
ROA	0.049	0.117	-0.386	-0.015	0.048	0.143	0.326
$MISP$	48.934	12.735	22.150	32.830	48.360	65.880	79.990
$IVOLA$	0.019	0.012	0.005	0.008	0.016	0.032	0.061
$HLSPREAD$	0.008	0.005	0.002	0.003	0.007	0.014	0.025
IO	0.561	0.261	0.026	0.184	0.578	0.903	1.000

Panel B: Correlation Table

	$SUSIR$	SR	DTC	SR_{IO}	$MBETA$	$SIZE$	BM	RET_RV	RET_MOM	INV	ROA	$MISP$	$IVOLA$	$HLSPREAD$	IO
$SUSIR$	1.00														
SR	0.22	1.00													
DTC	0.26	0.76	1.00												
SR_{IO}	0.26	0.91	0.79	1.00											
$MBETA$	0.00	0.17	0.05	0.15	1.00										
$SIZE$	-0.02	0.22	0.04	0.07	0.05	1.00									
BM	-0.03	-0.22	-0.13	-0.19	-0.03	-0.25	1.00								
RET_RV	0.02	-0.02	-0.03	-0.02	-0.01	0.06	0.03	1.00							
RET_MOM	0.01	-0.07	-0.09	-0.06	-0.02	0.08	0.01	0.01	1.00						
INV	0.04	0.03	0.00	0.05	0.03	0.02	-0.22	-0.02	-0.04	1.00					
ROA	0.00	-0.05	-0.11	-0.09	-0.07	0.16	-0.38	0.01	-0.01	0.38	1.00				
$MISP$	0.02	0.13	0.13	0.19	0.10	-0.20	0.12	-0.02	-0.27	0.42	-0.31	1.00			
$IVOLA$	0.03	0.12	-0.07	0.16	0.25	-0.34	-0.08	-0.01	-0.11	0.07	-0.07	0.17	1.00		
$HLSPREAD$	0.01	0.27	0.13	0.31	0.21	-0.26	-0.06	-0.09	-0.18	-0.02	-0.15	0.19	0.56	1.00	
IO	-0.02	0.55	0.22	0.24	0.12	0.43	-0.17	0.00	-0.03	-0.02	0.08	-0.08	-0.04	0.05	1.00

Table 2:
Performance of $SUSIR$ Sorted Portfolios.
This table reports performance of equal-weighted (Panel A) and value-weighted portfolios (Panel B) formed by monthly sorting stocks into deciles on $SUSIR$ measure. The lower part of each panel reports performance and t-statistics of the long-short strategy that is long in stocks with the 10% (30%) lowest values of $SUSIR$ and short in stocks with the the 10% (30%) highest values of $SUSIR$. The performance measures are raw returns (RawRet) and factor model alphas. Estimated models are CAPM (CAPM), Fama and French (1993) 3-factor model (FF3), Carhart (1997) 4-factor model (C4), Carhart (1997) 4-factor model augmented by Pastor and Stambaugh (2003) liquidity factor (C4+LIQ), Fama and French (2015) 5-factor model (FF5), Hou et al. (2014) q-factor model (HXZ4), Carhart (1997) 4-factor model augmented by quality-minus-junk and betting-against-beta factors (C4+Q+B) and Stambaugh and Yuan (2016) 3-factor model (UMO3).

Panel A: Equal-Weighted Portfolio									
Decile	RawRet	CAPM	FF3	C4	C4+LIQ	FF5	HXZ4	C4+Q+B	UMO3
1 (Long)	1.002	0.359	0.163	0.239	0.234	-0.034	0.077	-0.030	0.374
2	0.936	0.290	0.066	0.121	0.144	-0.139	-0.058	-0.194	0.259
3	0.875	0.233	0.025	0.079	0.097	-0.218	-0.104	-0.246	0.176
4	0.849	0.215	-0.007	0.032	0.047	-0.240	-0.165	-0.295	0.157
5	0.786	0.151	-0.057	-0.018	-0.019	-0.260	-0.192	-0.309	0.101
6	0.790	0.151	-0.046	-0.006	-0.005	-0.278	-0.189	-0.332	0.086
7	0.659	-0.002	-0.210	-0.172	-0.176	-0.424	-0.382	-0.452	-0.016
8	0.634	-0.031	-0.254	-0.208	-0.200	-0.446	-0.387	-0.470	-0.054
9	0.517	-0.149	-0.355	-0.276	-0.276	-0.530	-0.448	-0.541	-0.110
10 (Short)	0.572	-0.098	-0.310	-0.250	-0.242	-0.495	-0.428	-0.470	-0.065
1-10	0.430	0.458	0.473	0.489	0.475	0.461	0.505	0.440	0.439
	(5.287)	(5.498)	(5.451)	(5.703)	(5.419)	(4.882)	(5.396)	(4.074)	(4.664)
L 30% - H 30%	0.363	0.387	0.391	0.391	0.397	0.360	0.392	0.337	0.346
	(6.633)	(6.892)	(6.57)	(6.526)	(6.673)	(5.595)	(6.325)	(4.397)	(5.121)

Panel B: Value-Weighted Portfolio

Decile	RawRet	CAPM	FF3	C4	C4+LIQ	FF5	HXZ4	C4+Q+B	UMO3
1	0.862	0.278	0.246	0.220	0.196	0.163	0.134	0.123	0.250
2	0.830	0.249	0.200	0.249	0.237	0.107	0.133	0.095	0.263
3	0.640	0.048	-0.028	-0.051	-0.051	-0.151	-0.148	-0.156	-0.020
4	0.636	0.019	-0.015	-0.045	-0.037	-0.205	-0.253	-0.231	-0.061
5	0.598	0.004	-0.040	-0.017	-0.027	-0.142	-0.140	-0.172	0.004
6	0.729	0.124	0.079	0.104	0.110	-0.042	-0.009	-0.048	0.069
7	0.604	0.040	-0.016	-0.040	-0.052	-0.201	-0.249	-0.275	-0.054
8	0.356	-0.229	-0.330	-0.339	-0.334	-0.477	-0.562	-0.539	-0.293
9	0.580	-0.016	-0.085	-0.054	-0.041	-0.194	-0.197	-0.164	0.026
10	0.515	-0.073	-0.182	-0.147	-0.132	-0.310	-0.332	-0.332	-0.050
1-10	0.347	0.350	0.428	0.368	0.327	0.474	0.466	0.455	0.300
	(3.304)	(3.012)	(3.81)	(3.132)	(2.855)	(3.64)	(2.927)	(2.832)	(2.279)
L 30% - H 30%	0.293	0.297	0.337	0.313	0.287	0.358	0.393	0.353	0.267
	(3.582)	(3.186)	(3.814)	(3.538)	(3.357)	(3.57)	(3.244)	(3.297)	(2.84)

Table 3:
Regression of $SUSIR$ Long-Short Strategy on Other Short Interest-Based Long-Short Strategies

The returns of the $SUSIR$-based long-short strategy are regressed on the four factors from Carhart (1997) model and long-short strategies based on SR, DTC and SR_{IO}. All strategies are long in stocks from the lowest decile of the corresponding variable and are short in stocks from the highest decile of the corresponding variable. In Panel A (Panel B) returns of all strategies are calculated using an equal-weighting (value-weighting) procedure. The t-statistics are adjusted for autocorrelation following Newey and West (1987) with the lag of twelve months. ***, ** and * indicate significance at 1%, 5% and 10% levels, correspondingly.

Panel A: Equal-Weighted Strategy					
	(1)	(2)	(3)	(4)	(5)
$\_CONS$	0.489***	0.255***	0.330***	0.279***	0.242***
	(5.70)	(2.74)	(3.82)	(3.10)	(2.63)
LS_SR		0.267***			0.242***
		(4.76)			(3.33)
LS_DTC			0.183***		0.0330
			(3.02)		(0.56)
LS_SR_{IO}				0.214***	0.00699
				(4.19)	(0.13)
$MKTRF$	-0.0562**	0.0572**	-0.0331	0.00761	0.0527**
	(-2.23)	(2.22)	(-1.54)	(0.38)	(1.97)
SMB	0.00318	0.0489	0.00466	0.0526	0.0464
	(0.08)	(1.36)	(0.14)	(1.30)	(1.23)
HML	-0.0400	-0.0584**	-0.0330	-0.0489*	-0.0557*
	(-1.17)	(-1.99)	(-1.08)	(-1.68)	(-1.93)
UMD	-0.0179	-0.0438	-0.0236	-0.0356	-0.0430
	(-0.43)	(-1.35)	(-0.60)	(-0.89)	(-1.31)
N	404	404	404	404	404
Panel B: Value-Weighted Strategy					
	(1)	(2)	(3)	(4)	(5)
$\_CONS$	0.368***	0.288**	0.328***	0.251**	0.254**
	(3.13)	(2.36)	(2.69)	(2.11)	(2.02)
LS_SR		0.166**			0.0362
		(2.48)			(0.58)
LS_DTC			0.183***		0.100
			(2.62)		(1.55)
LS_SR_{IO}				0.201***	0.128**
				(2.86)	(2.00)
$MKTRF$	-0.0358	0.0319	-0.0352	0.0204	0.0152
	(-0.81)	(0.82)	(-0.85)	(0.52)	(0.38)
SMB	0.0379	0.0623	0.00938	0.0654	0.0452
	(0.81)	(1.15)	(0.21)	(1.25)	(0.82)
HML	-0.150**	-0.162**	-0.162***	-0.158**	-0.165***
	(-2.17)	(-2.50)	(-2.66)	(-2.54)	(-2.72)
UMD	0.0684	0.0602	0.0710	0.0538	0.0587
	(1.12)	(1.08)	(1.25)	(0.96)	(1.08)
N	404	404	404	404	404

Table 4:
Fama-MacBeth Regression of Stock Returns on $SUSIR$

This table reports results of the Fama-MacBeth regression of raw monthly stock returns on $SUSIR$ measure and other stock characteristics. All explanatory variables are normalized to have mean equal to zero and standard deviation equal to one. Additional control variables are market beta, log size, log book-to-market ratio, momentum and short-term reversal. The t-statistics are Newey and West (1987) adjusted. ***, ** and * indicate significance at 1%, 5% and 10% levels, correspondingly.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$SUSIR$	-0.114***	-0.0852***	-0.0928***	-0.0812***	-0.113***	-0.107***	-0.0998***	-0.0766***
	(-5.73)	(-4.55)	(-4.60)	(-3.97)	(-5.71)	(-5.61)	(-5.20)	(-4.04)
SR		-0.404***						0.168*
		(-4.77)						(1.70)
DTC			-0.169***					-0.0998**
			(-4.98)					(-2.28)
SR_{IO}				-0.234***				-0.168***
				(-6.55)				(-3.14)
INV					-0.161***			0.0280
					(-6.42)			(0.85)
ROA					0.134***			-0.0673
					(2.61)			(-1.20)
$MISP$						-0.274***		-0.222***
						(-6.64)		(-4.75)
$IVOLA$							-0.430***	-0.318***
							(-5.89)	(-4.95)
$MBETA$	-0.0208	0.00346	-0.0196	-0.00113	-0.0118	0.0154	0.0521	0.0636
	(-0.36)	(0.06)	(-0.34)	(-0.02)	(-0.21)	(0.27)	(0.93)	(1.17)
LN_SIZE	0.0276	0.00421	0.00166	-0.0368	0.00513	-0.0218	-0.102	-0.167***
	(0.43)	(0.07)	(0.03)	(-0.58)	(0.08)	(-0.35)	(-1.62)	(-2.66)
LN_BM	0.0958*	0.0843*	0.0905*	0.0707	0.0992**	0.113**	0.0524	0.0428
	(1.91)	(1.67)	(1.78)	(1.37)	(2.06)	(2.32)	(1.06)	(0.91)
RET_RV	-0.335***	-0.342***	-0.338***	-0.381***	-0.347***	-0.350***	-0.329***	-0.392***
	(-6.35)	(-6.46)	(-6.28)	(-6.93)	(-6.54)	(-6.58)	(-6.12)	(-6.96)
RET_MOM	0.249**	0.242**	0.240**	0.230*	0.246**	0.141	0.242**	0.124
	(2.09)	(2.02)	(2.03)	(1.88)	(2.05)	(1.23)	(2.14)	(1.04)
N	577088	577088	577056	475372	571201	577088	576894	470396
R^2	0.058	0.062	0.061	0.061	0.065	0.062	0.065	0.084

Table 5:
$SUSIR$, Biased Expectations and Fundamental News

This table shows estimation results for the panel regression of earnings surprises measures on the most recent $SUSIR$ and other control variables. All specifications include month-fixed effects. For the calculation of the first dependent variable, SUE^{PE}, past earnings per share (EPS) are used to estimate earnings surprises. SUE^{PE} is equal to the difference between current earnings per share and earnings per share reported 4 quarter ago divided by the standard deviation of this difference over past 8 quarters. For the calculation of the second dependent variable, SUE^{AF}, analyst forecasts are used as a proxy for the market expectations. SUE^{AF} is equal to EPS announced at month t net of the most recent mean analyst forecast divided by standard deviation of the most recent analyst forecasts. The third dependent variable, CAR, is the cumulative market-adjusted return over the earnings announcement window $[-1, 1]$. Analyst forecasts are obtained from IBES database and are required to be not older than 90 days. All earnings surprises measures are winsorized at 1% and 99% levels. All explanatory variables are normalized to have mean equal to zero and standard deviation equal to one. The standard errors are double-clustered by month and stock. ***, ** and * indicate significance at 1%, 5% and 10% levels, correspondingly. The sample period starts in May 1985 for SUE^{PE}, in May 1992 for SUE^{AF}, and in May 1980 for CAR.

	(1) SUE^{PE}	(2) SUE^{PE}	(3) SUE^{PE}	(4) SUE^{PE}	(1) SUE^{AF}	(2) SUE^{AF}	(3) SUE^{AF}	(4) SUE^{AF}	(1) CAR	(2) CAR	(3) CAR	(4) CAR
$SUSIR$	-0.0302***	-0.0366***	-0.0342***	-0.0254***	-0.0515***	-0.0579***	-0.0494***	-0.0419***	-0.0515***	-0.0513***	-0.0495***	-0.0364**
	(-3.51)	(-4.71)	(-4.42)	(-3.31)	(-3.61)	(-4.41)	(-3.85)	(-3.09)	(-3.08)	(-2.99)	(-2.91)	(-2.17)
$MISP$			-0.117***	-0.110***			-0.316***	-0.311***			-0.0931***	-0.0815***
			(-10.61)	(-9.94)			(-17.05)	(-16.92)			(-4.00)	(-3.59)
SR				-0.0539***				-0.0413**				-0.111***
				(-4.90)				(-2.05)				(-3.40)
$MBETA$		-0.0537***	-0.0393***	-0.0341***		0.0239	0.0603***	0.0639***		0.00423	0.0166	0.0276
		(-5.20)	(-3.88)	(-3.34)		(1.30)	(3.37)	(3.53)		(0.16)	(0.65)	(1.09)
LN_SIZE		0.156***	0.132***	0.124***		0.157***	0.0720***	0.0570**		0.0406*	0.0222	0.0114
		(10.20)	(8.98)	(8.58)		(5.47)	(2.62)	(2.03)		(1.70)	(0.95)	(0.48)
LN_BM		-0.111***	-0.0972***	-0.104***		-0.0592***	-0.0184	-0.0229		0.0273	0.0344	0.0232
		(-9.46)	(-8.42)	(-8.86)		(-2.90)	(-0.95)	(-1.19)		(1.17)	(1.47)	(1.00)
RET_RV		0.149***	0.148***	0.148***		0.268***	0.268***	0.268***		0.0167	0.0153	0.0153
		(9.08)	(9.14)	(9.34)		(11.63)	(11.95)	(12.10)		(0.59)	(0.54)	(0.54)
RET_MOM		0.411***	0.379***	0.380***		0.398***	0.318***	0.319***		0.0319	0.00625	0.00817
		(15.32)	(14.22)	(14.29)		(12.61)	(10.50)	(10.55)		(0.85)	(0.17)	(0.22)
$NUMEST$						0.00758*	0.0108***	0.0122***				
						(1.95)	(2.92)	(3.28)				
$STDEV$						-0.00403***	-0.00350***	-0.00325***				
						(-3.73)	(-4.02)	(-3.90)				
$FixedEffects$	Month	Month	Month	Month	Month	Month	Month	Month	Month	Month	Month	Month
N	145038	140366	140366	140366	124262	119874	119874	119874	195840	189153	189153	189153
R^2	0.028	0.080	0.084	0.084	0.013	0.031	0.038	0.038	0.007	0.007	0.007	0.007

Table 6:
Anomaly Returns on Earnings Announcement Days

The table reports the results of daily panel regression of raw stock returns on $SUSIR$ measure, earnings announcement period (EAP) dummy and their interaction. Column (1) includes day-fixed effects. Column (2) adds four lags of stock return, return squared and turnover ($Controls$), coefficients are not displayed. In Column (3) SR and $MISP$ and their interactions with EAP are included. In Column (4) day-fixed effects are replaced with more restrictive day-EAP fixed effects. In Column (5) return of CRSP value-weighted index (MKT), its interaction with $SUSIR$ and EAP and their triple interaction are added. The standard errors are clustered by month. ***, ** and * indicate significance at 1%, 5% and 10% levels, correspondingly.

	(1)	(2)	(3)	(4)	(5)
EAP	0.0633***	0.0611***	0.0612***		0.0460***
	(9.26)	(8.93)	(9.00)		(4.71)
$SUSIR$	-0.00441***	-0.00500***	-0.00361***	-0.00507***	-0.00410*
	(-3.79)	(-4.36)	(-3.21)	(-4.41)	(-1.90)
$SUSIR \times EAP$	-0.0174***	-0.0180***	-0.0141**	-0.0151***	-0.0172***
	(-2.99)	(-3.11)	(-2.47)	(-2.68)	(-2.70)
$MISP$			-0.0109***		
			(-5.29)		
$MISP \times EAP$			-0.0264***		
			(-3.84)		
SR			-0.0109***		
			(-3.43)		
$SR \times EAP$			-0.0292***		
			(-3.14)		
MKT					1.005***
					(105.04)
$MKT \times EAP$					0.00793
					(0.54)
$MKT \times SUSIR$					-0.00567
					(-1.27)
$MKT \times SUSIR \times EAP$					0.0204**
					(2.35)
$Controls$	None	Yes	Yes	Yes	Yes
$FixedEffects$	Day	Day	Day	Day*EAP	None
R^2	0.207	0.208	0.208	0.210	0.181
N	12552943	12537383	12537383	12537348	12537383

Table 7:
Analysis of Limits to Arbitrage

This table reports the estimation results for the Fama-MacBeth regression of monthly stock returns on $SUSIR$ measure and its interactions with variables that proxy for limits to arbitrage. These proxies are Corwin and Schultz (2012) spread ($HLSPREAD$), idiosyncratic volatility ($IVOLA$) and residual institutional ownership ratio (RIO). The interaction variables are sorted into quintiles. $SUSIR$ is normalized to have zero mean and unit standard deviation. It is interacted with the quintile dummies of the interaction variable. Dummies for the lowest quintile of the interaction variable is omitted and thus serves as a reference level. Additional control variables are market beta, log size, log book-to-market ratio, momentum and short-term reversal (coefficients are not displayed). The t-statistics are Newey and West (1987) adjusted. ***, ** and * indicate significance at 1%, 5% and 10% levels, correspondingly.

	$M = HLSPREAD$	$M = IVOLA$	$M = RIO$
	(1)	(2)	(3)
$SUSIR$	-0.0816**	-0.0352	-0.0893***
	(-2.49)	(-1.06)	(-2.84)
$SUSIR \times M_{Quintile=2}$	0.0339	-0.0618*	-0.0164
	(0.87)	(-1.74)	(-0.35)
$SUSIR \times M_{Quintile=3}$	-0.0287	-0.0837*	-0.0267
	(-0.57)	(-1.79)	(-0.58)
$SUSIR \times M_{Quintile=4}$	-0.0545	-0.0928*	-0.0687
	(-1.16)	(-1.87)	(-1.35)
$SUSIR \times M_{Quintile=5}$	-0.154***	-0.141**	-0.0209
	(-3.07)	(-2.38)	(-0.32)
$M_{Quintile=2}$	0.0348	0.0551	0.260***
	(0.75)	(0.92)	(4.77)
$M_{Quintile=3}$	0.0120	-0.00641	0.205***
	(0.26)	(-0.08)	(2.91)
$M_{Quintile=4}$	0.00578	-0.0927	0.254***
	(0.08)	(-0.91)	(3.10)
$M_{Quintile=5}$	-0.298**	-0.641***	0.189*
	(-2.55)	(-3.81)	(1.83)
$Controls$	Yes	Yes	Yes
R^2	0.0684	0.0719	0.0689
N	577088	576894	575995