HEC PARIS

Master in International Finance

RESEARCH PAPER

Academic Year 2025 - 2026

Short Interest as a Trading Signal: Predictive Power, Strategy Performance, and Implementation Frictions

Aya BEDHIAF

Vasara SILININKAITE

Under the supervision of Prof. Augustin LANDIER

President of the Jury: Prof. Alina ROSU

Keywords: short interest; stock returns; return predictability; long-short strategies; transaction costs; short-sale constraints

☒ *PUBLIC REPORT* ☐ *CONFIDENTIAL REPORT*

Abstract

This paper investigates the relation between short interest and future US stock returns, examining whether this signal can be translated into a profitable, implementable trading strategy. Using a dynamic universe of the top 2,000 US equities from 2007 to 2024, the short interest ratio is found to negatively predict future returns. A long-short portfolio strategy generates positive gross returns and significant factor-adjusted alphas. However, transaction costs reduce net returns, making most strategies difficult to implement profitably. Squeeze-filtered and long-only exclusion strategies provide alternative ways to exploit the signal through more applicable strategies, despite meaningful implementation frictions.

Keywords: Short interest; stock returns; return predictability; long-short strategies; transaction costs; short-sale constraints.

[1]Following HEC guidelines, we declare the application of Large Language Models (i.e. Claude and Chat GPT) to improve the writing style, grammar and spelling for clarity and proof-reading.

[2]We would like to extend our gratitude to Professor Augustin Landier for his supervision and insightful guidance throughout this project.

Contents

1 Introduction

Short selling is a critical mechanism in financial markets, enabling investors to express their views and correct existing asset prices. It is a practice mainly contained to highly sophisticated investors as shorting is a costly practice, that involves several frictions, such as search costs, borrow fees and others. Furthermore, investors are exposed to unlimited downside if asset prices continue to rise, further incentivizing only the most informed investors to participate in this side of financial markets. Given these characteristics, short positions are likely to reflect important information about the company, whether about its fundamentals, expected future performance and other factors – it becomes a factor which can explain future stock returns.

Existing literature support highly shorted stock underperformance; however, this signal is not sufficient to construct a tradable strategy. Despite initial papers, there currently exists a gap between the signal and introducing a tradable strategy based on findings of short interest predictability that the paper explores. The paper includes a sector analysis to further explore particularities of short interest that has not been previously explored in literature.

The paper will use US common stocks from 2006 till 2024 to gather information on short interest and the related factors. OLS and panel regressions are run to establish a relationship between short interest and future stock returns, followed by portfolio construction, controlling for variables and finally including transaction costs to test for the strategy's implementation.

This thesis will explore whether short interest is able to predict future returns, and whether this signal can be transformed into a strategy which is profitable and implementable. The specific trading strategy explored is shorting highly shorted stocks and long low-shorted stocks, as we assume short interest is informative regarding future stock returns – the more shorted a stock is, the greater underperformance we expect in the future. This strategy endures robustness tests, control variables and transaction costs to understand the main drivers behind the strategy and whether it is implementable in the context of these market conditions.

2 Literature review

2.1 Short interest and predictability

Short selling is a process within financial markets that enables a party to borrow an asset from its owner to sell it and buy it back at a further point in time. The party expects the price of the asset to decline and make a profit by buying back the asset at a lower price than it sold it, before returning it to the asset's owner. Short selling is an important market mechanism for speculative or arbitrage reasons: it is a way to express pessimistic views about the asset's price evolution or a way to correct its current valuation or price. Through this transaction, the market puts downward pressure on the asset through its sale. Furthermore, unlike purchasing assets, short selling has unlimited downside; hence, more informed market players tend to participate in these transactions. It is critical to understand who participates and how they use short selling as it can become an indicator of their views. If we assume the majority of short sellers are informed, increased shorting can indicate negative future returns, or if we assume it reflects excessive pessimism, then increased shorting can predict reversals or short squeezes. Bank of America reinforces that hedge funds and institutional investors are the main parties that short stocks and estimate up to 85% of short positions in stock markets are taken by hedge funds, which are extremely sophisticated investors (Ji, 2024).

Given the active and abundant participation of sophisticated investors, their investments can be used as a predictor of future market performance. Early academic papers, such as Branch (1976) and Hanna (1976), treated short interest as a market indicator, with subsequent research developing the theoretical link between short-sale constraints, overpricing, and subsequent underperformance, most notably in Jones and Lamont (2001). Even today academic literature finds short-selling activities have predictive power, such as in Akbas et al. (2017), and Kelley and Tetlock (2017) research.

2.2 Short-sale constraints and informed trading

In financial theory only a certain group of investors must be informed and have the ability to correct prices. In the case of shorting, these investors incorporate negative fundamentals, reverse over-optimistic views and others. These informed sellers can suppress the valuation and price in order to incorporate the true value of the asset, earning low future returns; hence, short selling can become an indicator of future market returns. However, in reality markets have many constraints and other limitations that prevent the reflection of assets' true valuations. As Grullon et al. (2015) find in their paper, short-selling regulatory constraints prevent information from being fully incorporate, especially the negative type that is reflected through shorting. Through using Regulation SHO, they observed that relaxing constraints affected stock prices negatively, causing valuations to decline; in turn financing decisions were also affected, seeing worsened conditions, especially amongst smaller firms. Duffie et al. (2002), and Jones and Lamont (2001) further show non-regulatory constraints, such as margin requirements, search frictions and short-squeeze risk prevent accurate reflection of asset valuations. Despite informed traders making up the majority of short sellers (Diether et al., 2008), these constraints can create such a burden and disincentive to trade, which may result in less downward pressure than there should be, causing overvaluation (Bekjarovski, 2018).

Hedge funds have benefited from being among the most informed in the industry and having the capacity to exploit market mispricing, bringing record profits over the past few years, especially recording their strongest gains since 2020 through the tech rally (Mourselas and Pollard, 2026). However, there are increasing questions whether these institutions have the upper hand. Through trading frenzies and meme stocks, discussions were being raised "about market efficiency, regulation and financial stability" (Plender, 2021). Retail investors squeezed the institutional investors; however, this occurred through price-insensitive investors, rather than being controlled by fundamentals. As Plender (2021) points out, this raises concerns for misallocation of capital, especially at that time when monetary policy was more relaxed, enabling suboptimal projects to be funded. These trends show the duality of markets: at times hedge funds are best positioned to exploit mispricing, but their hands may be tied by increasingly price-insensitive investors.

2.3 Early evidence on short interest indicators

Earliest papers used short interest to understand market sentiment and in turn predict returns (Hanna, 1976), but with time researchers added modifications to test for various constraints and scenarios to formally test the significance of short interest as a way to predict market returns. As Diether et al. (2008) summarize findings, that investors short firms with weak fundamentals, earning restatements, high accruals, analyst forecast revisions, and negative earnings surprises, reflecting investor competence in shorting. Nevertheless, despite the significant relationship between shorting and returns, recent papers started accounting for frictions affecting shorting, creating more formal tests (Asquith et al., 2005). Furthermore, researchers introduced different measures of short interest to predict future returns. For instance, in Boehmer et al.'s (2021) study they test for short interest ratio (shares on loan divided by shares outstanding), days-to-cover-ratio (shares on loan divided by daily trading volume), loans supply (shares available for borrowing divided by shares outstanding), utilization ratio (shares on loan divided by shares available for borrowing), demand outward shift (a dummy variable capturing an increase in shorting demand), among others. With each ratio a different economic implication is reflected, such as liquidity, cost and other factors that may influence shorting. Days-to-cover-ratio and utilization ratio were observed as the most powerful and robust ratios to predict future market returns.

Despite the insightful findings of all the papers, there are significant limitations that can limit the impact of their findings. Most notably earlier research papers (Hanna, 1976) fail to consider formal tests that would account for constraints that exist in financial markets, giving perhaps too optimistic data. Furthermore, others tend to rely on a small universe, using more historical data, which may no longer reflect current trends (Jones and Lamont, 2001), or the data comes from specialized data sets, complicating trading strategy construction (Diether et al., 2008). The majority of papers available also do not focus on constructing trading strategies, rather focusing on valuation distortions and explaining the information channel of the signal. Nevertheless, with time more recent papers bring new insights with more and latest data, constructing more robust results.

2.4 Cross-sectional evidence and trading-strategy implications

As academics across time demonstrated, there is a strong relationship between high short interest and lower subsequent returns, Akbas et al. (2017) demonstrate that this relationship is large and statistically significant, with the effect being greater for firms that are already likely to be targeted by short sellers. When building a trading strategy, there are two ways researchers have addressed it so far - through equal-weighted portfolios and those weighted by market capitalization. Asquith et al. (2005) indeed support that portfolios that short highly shorted stocks and long the less shorted stocks do perform overall, but this effect is more pronounced in equal-weighted portfolios rather than value-weighted portfolios. This is explained by smaller firms having a larger impact in an equal-weighted portfolio, where the effect can be more pronounced given the low institutional ownership and possibly increased high shorting demand. Nevertheless, it is difficult to effectively execute a value-weighted portfolio due to the constant rebalancing and the involved transaction costs that will reduce returns further.

Further studies expand their data by including daily short-selling data rather than monthly, allowing to better explore short-term trends. Diether et al.'s (2008) paper explores a strategy of buying stocks with low short-selling activity and shorting stocks with high short-selling activity, generating significant results. They specifically find that shorting follows stocks exhibiting positive returns as they try to target stocks experiencing overreaction and are likely overpriced temporarily. Informed sellers analyze the stock and disagree with the price given fundamentals and take the costly bet against the drive of the average trader, but the price often reverses, generating returns for short sellers.

Following these strategies, short interest is a powerful indicator for returns, but as Akbas et al.

(2017) note, it can predict other factors: negative earnings surprises, downward analyst forecast revision, negative public news and future unfavorable firm information. Firm fundamentals become an important driver of shorting, meaning information discovery is critical to driving the returns of shorting. Akbas et al.'s (2017) paper demonstrates that the shorting strategy is not simply based on following the crowd, but rather on short sellers discovering negative information before it is priced in by most traders.

2.5 Liquidity, constraints and implementation costs

Despite the significant returns exhibited by the above-mentioned trading strategies, the ability of traders to capture these returns are limited by trading constraints, such as liquidity, borrowing limitations, implementation costs and others.

Schultz (2024) in his analysis of short sale constraints over a century effectively demonstrates that short-selling frictions delay pricing corrections: stocks with higher borrowing fees result in lower future returns that persists for a long time. These higher fees can be mostly observed in smaller companies due to the limited institutional ownership, which are critical to lending shares to short sellers. As Asquith et al. (2005) find, stocks with highest shorting demand and lower institutional ownership drive the most constraints as supply and demand of stocks do not match, driving up implementation costs of mentioned trading strategies.

Furthermore, Saffi and Sigurdsson (2011) confirm that stocks with limited lending supply further drive up implementation costs – short sellers have to pay more to secure the share or may not even be able to find it, all which increase search costs. As a result, these types of stocks exhibit lower price efficiency. Liquidity constraints are explored to confirm the relationship between short-sale constraints and price efficiency: despite including liquidity controls (i.e. turnover, bid-ask spreads, market capitalization), lending supply still remains a critical factor for price efficiency.

Bekjarovski (2018) explicitly tests for transaction cost implementation and finds that they do not completely reduce the base-line strategy profitability: if the long-only side is exploited, the strategy remains profitable, and the direct short can be replaced with a synthetic short to improve earnings. In the case of a long-only portfolio, the extensive shorting costs can be avoided but then the trader cannot exploit the overpriced stocks.

Given these liquidity, borrowing fee and implementation cost limitations short sellers are unable to effectively participate in the market, resulting in stocks becoming overpriced (Jones and Lamont, 2001) – they exhibit high valuations initially but transition to lower returns over time. As Grullon et al.'s (2015) findings reveal, through regulation SHO and relaxing shorting constraints, price began to fall and likely to return to their fundamentals. On the other hand, predictive power improves with trading constraints – the most informed sellers are the most incentivized to participate in the market. However, these constraints also limit the ability to exploit trading strategies and capture these abnormal returns, so it may not create necessarily better results than in a less restricted environment.

2.6 Aggregate short interest and market timing

The literature focuses on two types of predictability – cross-sectional and aggregate. Cross-sectional studies focus on the performance of highly shorted stocks against those that are less-shorted, critical information for creating long-short portfolios. Aggregate studies focus on whether market-wide short interest predicts future market returns, important for timing trades. Rapach et al. (2016) construct a Short Interest Index from aggregate short interest. After adjusting for the long-run upward trend in short interest, the researchers reveal that higher aggregate short interest predicts lower future market returns, implying that short sellers have information about future aggregate cash-flow conditions. Gorbenko (2023) contributes to this evidence internationally and finds that aggregate short interest negatively predicts market returns in 24 of 32 countries, with stronger predictability during downturns and in markets with greater short-sale constraints. These studies imply that short interest can be used as a stock-selection signal, and as a market-timing variable.

2.7 Research gap and thesis positioning

Analyzing short interest and its predictability remains a difficult topic due to the constraints brought upon. Firstly, papers refer to different types of data with varying quality and frequency: Schultz (2024) used proxies for borrowing fees which may not reflect true values, Boehmer et al. (2021) use international data, but the availability and quality of data vary, and papers can refer to daily short interest data, which is not always easily accessible to the public. Furthermore, the strength of short-interest predictability also depends on the time period being used. With the existing papers there is a wide variety of time periods used in the universe, sometimes from a few years up to a century, providing a variety of conclusions. Also, some researchers rely on very old data despite the availability of current information, depicting an unrealistic image of the market today. Despite discussing trading strategies, implementation costs are not fully covered: Diether et al. (2008) consider a short-term strategy which does not consider the cost of frequent rebalancing, Akbas et al. (2017) focus on the relationship between short interest and returns but discuss the direct costs affecting the trading strategies in a simplified manner.

The diversity among countries' regulations and even the changes within a single country can make comparability difficult across time and across regions, as predictive power depends on country-level regulation and other conditions. The construction of a trading strategy also poses limitations: whether to create a value-weighted or equal-weighted portfolio can change results significantly (Asquith et al., 2005) but they also have practical implications when transforming it into a strategy. There is no uniform way to address this issue as all types of methodologies have their advantages and disadvantages, but they are important to consider when constructing one's own research.

Given the existing literature, their findings and remaining areas of exploration, this paper examines whether short interest still contains predictive information and how it can be translated to a trading strategy that accounts for implementation frictions.

Our paper contributes to the current literature by introducing an implementable trading strategy following the findings of raw return predictability. The short interest signal is tested through manipulating the portfolio through liquidity constraints, borrowing costs, rebalancing, sector exposures, and short-squeeze risk. Using a dynamic top 2,000 U.S. equity universe, we are able to examine short interest across multiple horizons, construct long-short portfolios, apply transaction-cost and borrow-cost assumptions, and introduce a short-squeeze filter. Through our extensive tests and alterations, we are able to demonstrate the relationship between short interest and returns as well as the economic significance of a strategy implementing the findings of the relationship following considerations of trading frictions.

3 Methodology

3.1 Data, Sample Construction, and Variable Definitions

The following section discusses the paper's data sources, variable construction, and data-cleaning procedures.

3.1.1 Data Sources

Stock return, price, trading volume, shares outstanding, exchange information and other firm-related factors are collected from the daily CRSP (Center for Research in Security Prices) stock file from 2005 to 2024. The chosen CRSP sample includes common shares listed on the NYSE, AMEX, and NASDAQ. This restriction excludes preferred shares, certificates, trusts, and other non-common equity instruments, making the sample more comparable across firms.

Short interest data is collected from Compustat's Security Short Interest database. Compustat's GICS sectors are used for sector classifications. Standard asset-pricing factors, including the market, size, value, momentum, and risk-free rate factors, are obtained from the Kenneth R. French Data Library and are merged at the monthly frequency.

3.1.2 CRSP Data Cleaning and Market Capitalization

In order to construct the required dataset, the initial data must be cleaned. Firstly, prices are converted to absolute values. Shares outstanding, prices, returns, and trading volume are converted to numeric format, and observations with missing or invalid values in key variables are excluded. Market capitalization is computed as follows:

$$ME_{i,t} = |P_{i,t}| \times SharesOutstanding_{i,t}, \tag{1}$$

The universe is defined at the firm level: market capitalization is first computed at the security level and then aggregated across securities under the same PERMCO.

3.1.3 Universe Construction

The investment universe is constructed dynamically through size-based ranking. At the end of each calendar year $T - 1$, firms are ranked by their total market capitalization observed in the company's last trading day of December. The universe for year T is then defined as the top 2,000 firms by firm-level market capitalization:

$$\mathcal{U}_T = \{f : Rank_{f,Dec(T-1)} \leq 2000\}. \tag{2}$$

This procedure ensures that only information available before the beginning of year T is used to construct the sample, thereby avoiding look-ahead bias. The universe defined for year T is applied to all daily observations throughout T, while the universe's composition can evolve annually.

Choosing a universe of 2,000 firms annually is a trade-off between market coverage and the ability to invest. A large universe provides cross-sectional variation in short interest and enables reliable portfolio sorts. Excluding the smallest firms reduces the impact of microcap stocks, and their extreme returns, limited liquidity, short-sale constraints, and data quality issues. The selected universe intends to approximate a liquid and institutionally investable U.S. equity sample. Nevertheless, the actual number of firms observed in T can be slightly lower than 2,000 because the sample undergoes CRSP exchange-code and share-code filters, valid price and shares-outstanding requirements, and company-level aggregation using PERMCO. Additionally, the firms in the universe at $T-1$ may not appear in year T due to delistings, mergers, acquisitions, or bankruptcies. Further mergers applied to short interest, sector, and future-return data can reduce the number of observations in specific empirical tests.

3.1.4 Firm-Level Representation

Firms with multiple securities are collapsed to a single security per firm-date. For each PERMCO-date, the security with the largest market capitalization is retained:

$$i^*_{f,t} = \arg\max_{i \in f} ME_{i,t}. \tag{3}$$

The following ensures that each firm has one observation per date and prevents firms with multiple share classes from receiving inaccurate weighting. The finalized security is then used for the construction of returns, trading volume, turnover, and the merge with short interest data.

3.1.5 Short Interest Variables

Short interest is measured as the number of shares sold short and not yet covered. Since short interest is reported at a lower frequency than daily CRSP data, specifically monthly, and more recently bi-weekly, it is merged onto the daily panel through a backward-looking method. Each daily observation is paired with the latest short interest report, treating short interest as a step function between reporting dates. To avoid carrying forward stale information for longer periods, observations for which the latest report is greater than 45 days are set to absent, ensuring that no further information is involved while reducing measurement error from outdated short interest reports. The paper employs a variety of short interest ratios to test different dimensions of short-interest predictability that is consistent with existing literature (Boehmer et al., 2021):

$$SIR_{i,t} = \frac{ShortInterest_{i,t}}{SharesOutstanding_{i,t}}. \tag{4}$$

The short interest ratio variable measures the fraction of shares outstanding that were sold. A higher value indicates greater short-selling activity relative to the firm's equity size.

$$\Delta SIR_{i,t} = SIR_{i,t} - SIR_{i,t-1}. \tag{5}$$

The monthly change in short interest ratio reflects changes in short positioning over time, which may capture new information, changing investor beliefs, hedging activity, or other trading motives.

$$DTCR_{i,t} = \frac{ShortInterest_{i,t}}{AverageDailyVolume_{i,t}}. \tag{6}$$

The days-to-cover ratio takes a fraction of short interest by average trading volume, measuring number of trading days required for short sellers to cover their positions under normal volume conditions. The ratio reflects the relation between short interest and the stock's liquidity (Boehmer et al., 2021).

3.1.6 Monthly Aggregation

The empirical analysis is conducted at the monthly frequency. Daily returns are compounded for each calendar month to construct monthly stock returns:

$$R^M_{i,t} = \prod_{d \in t}(1 + r_{i,d}) - 1. \tag{7}$$

Short interest, SIR, price, shares outstanding, market capitalization, and firm size are based on the last available observation in each month. Trading volume is averaged over the month, turnover is defined as trading volume divided by shares outstanding. These alterations create a monthly firm-level panel for predictive regressions and portfolio formation. Future returns are constructed over 1-, 3-, 6-, and 12-month horizons. For each firm-month observation, future returns are compounded through monthly returns:

$$R_{i,t+h} = \prod_{s=1}^{h}(1 + R_{i,t+s}) - 1. \tag{8}$$

where $h \in \{1, 3, 6, 12\}$. These future returns are used as the dependent variables in the predictive regressions and as the holding-period returns in the portfolio tests.

3.1.7 Control Variables and Final Panel

The analysis includes firm-level controls that could affect future returns independently of short interest. Firm size is measured as the natural logarithm of market capitalization, turnover reveals trading intensity, and average daily volume measures liquidity. Sector classifications are used to examine short interest predictability differences across sectors and to make sector-neutral strategies. Market variables, including Fama–French factors and the momentum factor, are merged at the monthly frequency to test strategy robustness and factor-adjusted performance tests.

The main explanatory variables and future return variables are winsorized at the 1st and 99th percentiles before running regressions to reduce the impact of extreme observations. Variables with missing or non-finite values, which are required for a given test, are excluded from it, reflecting differences in data availability, listing histories, short interest coverage, and return-horizon requirements.

The monthly panel contains 442,962 firm-month observations and 4,761 unique firms over the time frame of January 2006 to December 2024. The number of observations for each specification depends on the return horizon, the availability of lagged variables, the use of non-overlapping samples, and the variables needed. The constructed dataset will be used for return predictability tests and portfolio-based trading strategy analysis. The panel enables baseline regressions of future stock returns on short interest variables. The inclusion of short interest, liquidity, sector, market, and factor variables helps to evaluate a strategy's drivers, and transaction costs evaluate the strategy's implementation.

3.2 Empirical Design

The following section will explain the paper's methodology and the results found using the data sources described above. The subsections are organized according to the different empirical tests.

3.2.1 Baseline and Sector-Level Predictive Regressions

The first regression tests for a systematic relation between short selling at month t and stock performance over horizons of 1, 3, 6, and 12 months. The following regressions are applied in existing literature of Boehmer et al. (2021) and Diether et al. (2008). This regression analysis is the foundation for the portfolio strategies discussed later in the thesis. The following specification is estimated using pooled multivariate OLS:

$$R_{i,t+h} = \alpha + \beta_1 SIR_{i,t} + \beta_2 \Delta SIR_{i,t} + \beta_3 DTCR_{i,t} + \delta' X_{i,t} + \varepsilon_{i,t+h}, \tag{9}$$

where $R_{i,t+h}$ denotes stock i's cumulative future return from month $t+1$ to month $t+h$. The main explanatory variables are $SIR_{i,t}$, $\Delta SIR_{i,t}$, and $DTCR_{i,t}$. The vector $X_{i,t}$ includes firm-level controls: size, turnover, and average daily volume. Standard errors are clustered monthly.

The pooled multivariate OLS results (Figure 1) demonstrate initial evidence of a negative relation between short interest and future returns: SIR's coefficient is negative at all horizons. 1- and 3-month horizon observations are significant at 2% and 1% level respectively, with coefficients of -0.0535 and -0.2227, implying that the predictive power of short interest is strongest at shorter horizons. Only ΔSIR is significant at the 1-month horizon (10% level), while DTCR is significant at the 12-month horizon (5% level), with coefficients of 0.1488 and -0.0052, respectively. The explanatory power of pooled regressions is limited, given noisy future returns, which are difficult to explain through observable characteristics.

The following is a two-way fixed-effects panel regression:

$$R_{i,t+h} = \alpha_i + \gamma_t + \beta_1 SIR_{i,t} + \beta_2 \Delta SIR_{i,t} + \beta_3 DTCR_{i,t} + \delta' X_{i,t} + \varepsilon_{i,t+h} \tag{10}$$

where α_i denotes firm-fixed effects and γ_t denotes time-fixed effects. Firm-fixed effects control for persistent differences across firms and time-fixed effects for market-wide shocks affecting all firms in a month. This regression identifies the predictive relation from within-firm variation in short interest after removing common time effects. Standard errors are clustered by both firm and time.

This regression's results are statistically stronger than the pooled OLS results (Figure 2). SIR's coefficient is negative and statistically significant at every horizon, implying that increased short interest precedes lower future returns. In the fixed-effects regression, an increase in SIR by 1 predicts a decline in future returns by 0.0549 the next month, a cumulative decline by 0.2261 over the next 3 months, a decline by 0.3179 over the next 6 months, and a decline by 0.571 over the next year. The magnitude increases with the horizon, despite fewer non-overlapping observations over horizons.

This panel regression, along with the previous OLS regression, demonstrates that firm size is negative and statistically significant across all horizons (Figure 1 & 2) but turnover and average daily volume are not statistically significant as in the previous regressions (Figure1). This implies the negative relation between short interest and future returns is not explained by the liquidity controls considered in the model.

The other short interest ratios are less informative. ΔSIR's coefficient is positive and statistically significant only at the 1-month horizon at the 1% level, with a coefficient of 0.0597 (Figure 2). DTCR is not statistically significant in any of the regressions (Figure 1 & 2). These results suggest SIR is the most robust predictive variable, with the rest providing weaker and less consistent direct predictive power.

For horizons longer than 1 month non-overlapping observations are used for each firm. This reduces mechanical dependence across return windows because the same monthly stock return is not constantly included in overlapping future-return calculations.

Furthermore, univariate panel regressions are estimated by individually testing each explanatory variable with firm and time fixed effects (Grullon et al., 2015): SIR is no longer statistically significant (Figure 3), but size remains negative and significant across all horizons. DTCR is significant at the 12-month horizon at 1% level, with a coefficient of 0.003. Turnover becomes significant at 12-month horizon, and average daily volume at 3 and 6-month horizons.

The regression analysis is repeated at the sector level to examine the pattern of short interest and future returns across sectors – short interest may reflect different information across sectors. The sector-level analysis is implemented through univariate panel regressions. For each sector, we estimate univariate panel regressions of future returns on each explanatory variable separately. These regressions are used to compare the strength of the short-interest signal across industries. These regressions include time and firm fixed effects and cluster standard errors at the firm level.

There is substantial heterogeneity across sectors (Figure 4). SIR's coefficient carries different signs, and only a few industries in certain horizons show statistically significant result, suggesting the predictive power of short interest is not evenly distributed. In Figure 5, we can observe the sector-level panel over 3-month horizon. SIR's coefficient is negative in two of the eleven sectors and statistically significant among the following (Figure 4): information technology (-0.1193) and healthcare (-0.0838). For the 1-month horizon, healthcare carries a statistically significant coefficient at 5% (-0.0352) and information technology at 1% (-0.0413). At the 6-month horizon some predictive power remains, in information technology (-0.2131) and now consumer staples (0.2807) with a positive coefficient. At the 12-month horizon the sectors with predictive power change: energy (0.7506) at 5% level and information technology (-0.4022) at 10% level.

Information Technology demonstrates the most consistent results: the SIR coefficient is negative and statistically significant at all horizons, suggesting that short interest contains persistent information about following underperformance. Other sectors show strong evidence at fewer horizons. Figure 5 showcases the varied distribution of SIR across industry, reinforcing the importance of accounting for sector composition when moving from regressions to portfolio construction.

These results motivate the sector-neutral portfolio tests later in the thesis. If short interest is structurally higher or more informative in certain sectors, a raw short-interest sort may unintentionally capture sector exposure. Testing the strategy within sectors therefore helps determine whether the portfolio performance reflects stock-level short-interest information rather than broad industry differences.

3.2.2 Baseline Long-Short Portfolio Construction and Robustness Tests

Following the results of the predictive regressions, the short-interest signal can be used to construct a portfolio strategy which Diether et al. (2008) apply in their study. Monthly, at each rebalancing date t, stocks in the sample are ranked according to their $SIR_{i,t}$ and assigned to deciles: decile 1 contains low SIR stocks, while decile 10 contains high SIR stocks. Portfolio returns are computed as equal-weighted decile averages. For each horizon $h \in \{1, 3, 6, 12\}$, each decile's return is computed using the future return variable $R_{i,t+h}$:

$$R^D_{k,t,h} = \frac{1}{N_{k,t}} \sum_{i \in D_{k,t}} R_{i,t+h}, \tag{12}$$

The baseline strategy goes long decile 1 and short decile 10:

$$LS_{t,h} = R^D_{1,t,h} - R^D_{10,t,h}, \tag{11}$$

where $R^D_{1,t,h}$ is the return of the lowest SIR decile and $R^D_{10,t,h}$ is the return of the highest SIR decile. A positive value of $LS_{t,h}$ implies that low SIR stocks outperform high SIR stocks.

The long-short strategy generates positive average returns at all horizons (Figure 6). The average monthly return is 0.73% at the 1-month horizon, 1.85% at the 3-month horizon, 2.7% at the 6-month horizon, and 2.8% at the 12-month horizon. However, the long-horizon result is not statistically significant after accounting for autocorrelation. Figure 7 continues analysis on the 1-month returns, showing lower returns under the long-short strategy than the market benchmark; however, that is accompanied by a higher sharpe ratio for the long-short strategy. Figures 8–10 show an overall increasing trend for cumulative returns for the baseline long-short strategy; however, there is a very big dip and further rapid recovery around the beginning of 2020. Figure 11 shows a less clear trend for the 12-month horizon results.

The portfolio sorts overall confirm the previous regression findings: stocks with low short interest outperform stocks with high short interest, with the effect most pronounced at short-to-medium horizons. For horizons longer than one month, non-overlapping observations are used when plotting cumulative performance and evaluating statistical significance.

After constructing the baseline long-short strategy, it is critical to evaluate whether it remains meaningful under more realistic implementation constraints. The baseline strategy is a clean test of the signal but is unlikely to be fully implementable: it relies on the two most extreme portfolios, requires frequent rebalancing, and involves shorting the highest SIR stocks, hence exposure to crowded-short risk.

Before applying explicit transaction costs, several adjusted versions of the baseline strategy are tested. The following discussed modifications are commonly observed in literature, as in Asquith et al.'s (2005) paper. The first modification keeps the decile structure but reduces rebalancing frequency from monthly to quarterly, testing whether the signal remains significant under less frequent trading. The second modification replaces deciles with quintiles: quintile portfolios create broader samples, reduce ranking sensitivity to small month-to-month changes, and provide a less concentrated implementation of the signal. The third modification excludes the most-shorted decile before portfolio formation, as these stocks are likely to exhibit higher transaction and search costs. The remaining stocks are then sorted into quintiles by SIR: the lowest-SIR quintile is bought and the highest-SIR quintile is shorted. This modification avoids relying on the most crowded short positions while preserving exposure to the short-interest signal.

Figure 12 reports annualized returns for comparability. The baseline monthly decile strategy generates an annualized return of 9.17% at the 1-month horizon and 7.63% at the 3-month horizon, both significant at the 1% level. The signal remains positive under quarterly rebalancing, with an annualized return of 13.13% at the 1-month horizon and 8.44% at the 3-month horizon, both significant at the 1% level, suggesting that the signal is not driven solely by monthly rebalancing.

The quintile strategy provides a less concentrated sorting while preserving much of the baseline performance. The monthly quintile strategy earns annualized returns of 7.68% at the 1-month horizon and 6.42% at the 3-month horizon, both significant at the 1% level. Under quarterly rebalancing, the quintile strategy earns annualized returns of 12.09% at the 1-month horizon and 6.63% at the 3-month horizon, significant at the 1% and 2% levels respectively. The short-interest signal is not concentrated within deciles and survives broader groups.

Excluding the most-shorted decile and shorting the remaining stocks within quintiles maintains positive results but leads to a decline in performance relative to the benchmark. This strategy earns annualized returns of 4.46% at the 1-month horizon and 3.44% at the 3-month horizon, significant at the 5% level. Under quarterly rebalancing, it earns annualized returns of 8.18% at the 1-month horizon and 3.36% at the 3-month horizon. The 1-month result remains statistically significant at the 2% level, while the 3-month result is statistically weaker. The most shorted stocks therefore contribute significantly to baseline performance, although the signal does not disappear entirely when they are excluded.

Turnover confirms the benefit of broader and less frequently rebalanced strategies: the baseline monthly decile strategy has a monthly turnover of 0.42, while quarterly rebalancing reduces it to 0.21, and the monthly quintile strategy has a monthly turnover of 0.30, and its quarterly version reduces it to 0.16. When excluding the most-shorted decile, monthly turnover is 0.40 under monthly rebalancing and 0.22 under quarterly rebalancing. Overall, quarterly rebalancing materially reduces trading intensity, while quintile portfolios provide a less concentrated and lower-turnover alternative to the baseline strategy.

3.2.3 Factor-Adjusted Performance

It is critical to analyze the contribution of standard equity risk factors in the long-short returns, as the short-interest strategy may appear attractive because it captures common sources of return, such as market risk, size, value, or momentum (Diether et al., 2008). The previous long-short portfolio returns are therefore regressed on the market factor, the Fama–French three factors, and the Carhart four-factor model.

The factor-adjusted regressions are estimated using the long-short return series from the previous section, before transaction costs, in order to isolate the independent return information contained in the short-interest signal prior to implementation frictions. The monthly quintile strategy is employed, as this portfolio is less concentrated than the decile strategy, has lower turnover, and still preserves most of the return predictability as observed in Figure 13. For each holding horizon $h \in \{1, 3, 6, 12\}$, the following models are estimated:

$$LS_{t,h} = \alpha_h + \beta_h \,\mathrm{MKT}_{t,h} + \varepsilon_{t,h}, \tag{14}$$

$$LS_{t,h} = \alpha_h + \beta_h \,\mathrm{MKT}_{t,h} + s_h \,\mathrm{SMB}_{t,h} + v_h \,\mathrm{HML}_{t,h} + \varepsilon_{t,h}, \tag{15}$$

$$LS_{t,h} = \alpha_h + \beta_h \,\mathrm{MKT}_{t,h} + s_h \,\mathrm{SMB}_{t,h} + v_h \,\mathrm{HML}_{t,h} + m_h \,\mathrm{MOM}_{t,h} + \varepsilon_{t,h}. \tag{16}$$

The first model is the capital asset pricing model (CAPM), the second is the Fama–French three-factor (FF3) model, and the third is the Carhart four-factor model (FF4). $LS_{t,h}$ is the return of the strategy. Since the strategy is a self-financing long-short portfolio, the dependent variable is interpreted as a spread return. Factor returns are compounded over the corresponding holding horizon, and Newey–West standard errors are used with lags equal to the holding horizon.

Observing Figure 13 and Figure 14, for the monthly quintile strategy, annualized alphas are positive and statistically significant across all horizons and all factor models. At the 1-month horizon, the annualized alpha is 11.1% under the CAPM, 9.76% under FF3 model, and 9.36%

under the FF4 model, all significant at the 1% level, indicating that the short-interest strategy is not driven by market exposure. The results remain significant at the 3-month horizon: the annualized alpha is 9.55% in the CAPM, 8.13% in the FF3 model, and 7.83% in the FF4 model, all significant at the 1% level. At the 6-month horizon, the annualized alphas remain positive and statistically significant. At the 12-month horizon, the annualized alpha declines but remains significantly positive

These results show that the short-interest signal is not fully explained by standard equity risk factors, generating significant positive abnormal returns after controlling for market, size, value, and momentum exposures. The alpha is strongest at the 1- to 3-month horizons, consistent with the earlier portfolio-sorting results.

3.2.4 Conditional Factor Sorts

To examine where the return predictability is concentrated, conditional portfolio sorts are considered, a common methodology to control for other variables in literature (Zhu et al., 2019). At each month, stocks are first sorted into quintiles based on FF4 firm characteristics: size, momentum, market beta, and liquidity. Within each characteristic quintile, stocks are then sorted into short-interest quintiles based on $SIR_{i,t}$. The long-short return is computed as the equally-weighted return of the lowest-SIR quintile minus the highest-SIR quintile within each characteristic group. One-month-ahead winsorized returns are used, and all portfolios are constructed before transaction costs.

Observing Figure 15 and Figures 16–19, the short-interest effect is strongest among small firms. In the lowest size quintile, the annualized return is 15.32% and significant at the 1% level. Returns decline as firm size increases, reaching only 3.23% in the highest size quintile, which is no longer statistically significant. Figure 20 demonstrates significantly greater cumulative returns for the quintile with the smallest stocks, while other quintile returns are more concentrated. This suggests that short-interest predictability is concentrated among smaller firms, where limits to arbitrage, information frictions, and trading constraints are more severe.

Within momentum sorts, the short-interest strategy performs strongly among past losers. The annualized return in the lowest momentum quintile is 14.62% and significant at the 1% level, but the effect weakens and becomes insignificant among past winners. Figure 21 shows the greatest cumulative returns for the past losers and second quintile for smallest momentum. High short interest is therefore more informative among stocks that have recently underperformed, consistent with short sellers identifying firms with persistent negative information or financial distress.

For market beta sorts, the short-interest effect is present across all quintiles but is strongest among the extremes. In the highest beta quintile, the annualized return reaches 9.23% and is significant at the 1% level, whereas in the lowest-beta quintile, returns are 6.53% and significant at the 3% level. Figure 22 shows how cumulative returns produced roughly similar results up until 2023, where high beta stocks started producing greater returns relative to other quintiles. The effect is weaker in the middle beta quintiles, suggesting that short interest is particularly informative at the extremes of systematic risk exposure.

For liquidity sorts, the signal is significant across most quintiles. Annualized returns are positive across all groups and are greatest in the second and third liquidity quintiles, at 11.3% and 9.48%, respectively, significant at the 1% level. The highest-liquidity quintile exhibits lower returns that are not statistically significant. Figure 23 supports this, with most liquid stocks producing the most lagging cumulative returns among the quintiles. Overall, short-interest predictability is strongest outside the most liquid stocks, but is not exclusively driven by illiquid segments.

3.2.5 Sector-Neutral Portfolio Sorts

For additional robustness, we examine short-interest signal's survival after controlling for sector, given the slight heterogeneous results in the previous sector-level predictive regressions. The baseline portfolio sort ranks all the stocks, allowing the long-short strategy to capture sector-level differences if shorting is concentrated in industries. Stocks are sorted into SIR quintiles for each sector, longing the lowest-SIR quintile and shorting the highest-SIR quintile. Returns are equal-weighted and computed before transaction costs.

Observing Figure 24, the sector-neutral strategy generates positive average returns at all horizons: at the 1-, 3- and 6-month horizon, the strategy has an annualized return of 6.98%, 6.15% and 5.01% respectively, at the 1% level. The 12-month horizon returns 3.15% at a 5% level. The short-interest signal is not only driven by sector composition: low-SIR stocks continue to outperform high-SIR stocks. This supports that short interest includes firm-level information about future returns, rather than reflecting overall differences across industries.

3.2.6 Transaction Costs and Limitations

Following these robustness checks, explicit transaction costs are applied through a realized-turnover framework (Muravyev et al., 2017). For each portfolio, long and short legs' weights are tracked separately, and trading costs are computed from actual change in stock-level weights between rebalancing dates:

$$Turnover_t = \frac{1}{2} \sum_i |w_{i,t} - w_{i,t-1}|. \tag{10}$$

Trading costs are then calculated separately for the long and short legs:

$$TradingCost_t = Turnover_t^{Long} \times CostRate^{Long} + Turnover_t^{Short} \times CostRate^{Short}. \tag{11}$$

The baseline trading-cost assumptions are 25 basis points for the long leg and 50 basis points for the short leg, reflecting the greater cost of implementing the short position. Borrow costs are applied to the short leg and scaled by the holding horizon:

$$BorrowCost_{t,h} = BorrowFee \times \frac{h}{12}. \tag{12}$$

Given stock-level borrow-fee data is not available, portfolio-specific borrow-fee assumptions are applied. The baseline decile strategy assumes 5% annual borrow fees, 3% for the quintile strategy, and 1% for the D10-excluded quintile strategy, reflecting that the baseline decile strategy shorts the most crowded stocks, and excluding D10 strategy avoids expensive stock borrowing. A more conservative stress calibration uses borrow fees of 10%, 5%, and 3%, respectively.

$$NetReturn_{t,h} = GrossReturn_{t,h} - TradingCost_t - BorrowCost_{t,h}. \tag{13}$$

Observing Figure 25, implementation costs absorb a substantial part of the raw long-short returns: at the 3-month horizon, the baseline monthly decile strategy falls from a return of 1.85% to 0.47% under the main considerations, and the cost-adjusted return is no longer statistically significant. The quintile strategies outperform decile strategies after costs: the 3-month horizon exhibits the highest monthly quintile return at 0.72% but it is no longer statistically significant. The quarterly quintile strategy produces a similar net return with lower monthly-equivalent turnover, which makes it the most realistic implementation candidate among the tested strategies. By excluding D10, we have a more conservative portfolio; however, it reduces the strength of the signal statistically: avoiding the most crowded shorts reduces borrowing-cost, but removes important predictive power from the short leg.

The stress calibration further weakens the results as under higher borrow-fee assumptions, the baseline decile strategy returns become negative, while the quintile and exclude-D10 strategies are close to zero, showing the strategy is highly sensitive to borrow-cost assumptions and is not robust even under light shorting-cost scenarios, though almost all results become statistically insignificant (Figure 26). Following Figure 27 and 28, we observe returns decline rapidly with an increase in borrowing fees, especially among strategies that short highest-SIR stocks. Turnover is also critical to consider as it varies significantly across strategies (Figure 29) and therefore affects the sensitivity of the strategy to borrow fees. The transaction-cost analysis results in a cautious interpretation: the short interest implementable profitability is much weaker, usually insignificant, once trading costs are considered.

3.2.7 Short-Squeeze Risk and Squeeze-Filtered Portfolios

Stocks with high short interest, high days-to-cover, and strong recent returns may be exposed to rapid price increases if short sellers are forced to cover their positions, known as a short squeeze (Horstmeyer et al., 2023). This risk is not reflected by average borrow-fee assumptions: the possibility of large adverse short-leg returns during crowded short episodes is ignored. Hence, a squeeze-risk indicator using three conditions is constructed: a stock is classified as squeeze-risk if it is in the top quintile of SIR, the top quintile of DTCR, and the top quintile of past one-month returns. 2.56% of stock-month observations are classified as squeeze risk. Within the high-SIR leg, the average share classified as squeeze-risk is 11.94%, indicating that squeeze-prone stocks are a significant part of the short portfolio.

Two squeeze-filtered strategies are created: first, a monthly squeeze-filtered quintile strategy, buying the lowest-SIR quintile and shorting the highest-SIR quintile after removing stocks classified as squeeze-risk from the short leg, second, applies the same method with quarterly rebalancing. The squeeze-filtered strategies perform better than the other cost-adjusted strategies previously tested: at 1-, 3- and 6-month horizons the squeeze-filtered quintile monthly long-short strategy produces net returns of 4.80%, 4.67% and 3.92% respectively, at 10%, 5% and 10% level significance, whereas at 1- and 3-month horizons the quarterly equivalent strategy has net returns of 8.79% and 4.74%, both significant at the 10% level. The rest of the long-short strategies are not significant following transaction costs (Figure 31). Filtering out squeeze-risk stocks improves the predictive power and implementation of short-interest. The squeeze-filtered strategy remains exposed to the high short-interest anomaly while avoiding stocks that seem most vulnerable to short squeezes.

Figures 34 and Figure 35 show a positive trend in gross and net cumulative returns for squeeze-filtered quintile monthly and quarterly strategies at the 3-month horizon; however, there is a growing difference in the monthly returns, most likely attributed to higher rebalancing costs.

3.2.8 Long-Only Equal-Weighted Exclusion Portfolios

Given the elevated costs and restrictions of shorting in our strategies, a portfolio with only the long leg is tested to see if information on short interest can be used without using the short mechanism. A similar long-only implementation is examined by Bekjarovski (2018) paper. Firstly, a monthly equal-weighted benchmark portfolio is created with all the shares, followed by three long-only portfolios: stocks in the top 20% of the SIR distribution are excluded, in the second stocks both in the top 20% of SIR and in the bottom 20% of past momentum are excluded, in the third stocks in top 20% of SIR or in the bottom 20% of past momentum are excluded. SIR and momentum are the main criteria for portfolio construction given they are among the strongest signals from the conditional sorts.

The long-only portfolio's monthly return is the equal-weighted average return of considered stocks:

$$R_{t+1}^{LO} = \frac{1}{N_t} \sum_{i \in P_t} R_{i,t+1}, \tag{14}$$

where P_t are the stocks complying with the strategy's criteria.

Transaction costs are calculated through realized turnover, and no borrow fees or short-side trading costs are considered:

$$NetReturn_{t+1}^{LO} = R_{t+1}^{LO} - Turnover_t \times CostRate^{Long}. \tag{15}$$

Observing Figure 34, the only strategy from the above mentioned that beat an equal-weighted benchmark index was the one excluding high SIR and low momentum stocks: net returns were consistently greater across all horizons and all significant at the 1% level. Among the remaining strategies, going long stocks that excluded top 20% SIR produced higher net returns compared to the strategy that excluded either high SIR stocks or low momentum stocks. However, the best performing strategy also has the highest Sharpe ratio as observed in Figure 39, which implies higher risk taken for the higher returns.

4 Analysis

4.1 Interpreting short interest as a trading signal and its implementation credibility

Short interest is a critical metric that can be used to determine a stock's profitability. Following results from baseline regressions (Figure 1 & 3), SIR produced statistically significant results when predicting future returns. Changes in SIR and DTCR did not produce statistically meaningful results, which supports short interest as a variable that compounds negative information or persistent pessimistic outlook that Akbas et al.'s (2017) research supports. This can explain why the sign of SIR's changes – the variable reflects short-term outlook, hedging, which can rapidly change. At this stage DTCR is less informative as a direct predictor of returns, which may be as a result of the monthly frequency, but liquidity does maintain a critical role in explaining the short interest signal. DTCR's significance contradicts Boehmer et al.'s (2021) findings: DTCR was among the most significant variables; however, our universe is constrained to a more investable sample and does not consider the global context. Even if the initial OLS and panel regressions have low explanatory power the results are statistically meaningful – the former is a result of noisy future returns at stock-level. The regressions run above overall support the hypothesis that high short interest predicts lower future stock returns, especially the two-way fixed-effects panel. SIR's significance in the multivariate panel regressions but not in the univariate suggests its predictability is better observed after controlling for firm characteristics and other short-interest-related variables. These results justify SIR's usage as the main sorting variable in the later portfolio analysis.

Using this significance, we test for economic tradability through portfolio construction, with varied controls. In the baseline case (Figure 6) we observe positive and statistically significant results among the short- and medium-term horizons, consistent with our initial hypothesis. Figure 7 further expands on the risk profile of the strategy: the baseline long-short strategy has lower annualized return than the market benchmark, but volatility is also lower, accompanied by a higher Sharpe ratio, a smaller maximum drawdown, and almost zero correlation with the market. The greater Sharpe ratio suggest a more attractive risk-return profile. This strategy reveals a cross-sectional return source that could diversify existing strategies. Furthermore, modifying the portfolio construction, quintile strategy, quarterly rebalancing, excluding D10, weakens the signal but it remains positive and significant, suggesting crowded short positions do not alone explain predictability and lower turnover through quarterly rebalancing improves the strategy's implementation. The factor-adjusted strategy through persistent alpha shows that returns are not entirely justified by market, size, value or momentum. Figure 15 and Figures 16–19 further expand the concentration of the signal amongst certain firm characteristics: the short signal is stronger among smaller-firm stocks, past losers, high-beta stocks. These findings align with the assumption that information asymmetry, distress, uncertainty and limits to arbitrage can amplify the predictive power of the short interest. These conditions promote only the most sophisticated traders to participate as it becomes too costly to participate without sufficient information and resources. Lastly, observing Figure 24, which controls for sector differences, positive sector-neutral returns imply that the strategy is not focused on shorting high-SIR sectors and buying low-SIR sectors, expanding its application market wide.

Despite the strong signals from all the modifications and constructions, its implementation becomes unrealistic when considering the borrowing fees as all strategies become insignificant at the 10% level after controlling for autocorrelation (Figure 25). This highlights the difficulty of actually monetizing the long-short strategy due to its unique limitations, specifically regarding borrowing and search fees, which are much greater on the short leg rather than long leg. However, the squeeze-filtered strategies exhibit greater significance and positive results (Figure 31), showing that with certain modifications, we are still able to take advantage of the short interest signal analyzed previously in the section and limit exposure to stocks at risk of short squeezes. However, the strategy can also limit its exposure to the shorting mechanism but still take advantage of the shorting signal: going long a portfolio that excludes highly shorted SIR and low momentum stocks. As observed in Figure 38 and 39, returns may be higher for this strategy compared with the equity benchmark index, and sharpe ratio is ever so slightly greater than the benchmark in the short run, indicating opportunities for portfolio diversification. This has important implications for future portfolio constructions – despite the short leg being difficult to implement because of borrowing and search costs, we are still able to construct a portfolio without short exposure, though it may not be the exact baseline long-short portfolio discussed before.

4.2 Sector heterogeneity and information content

Short interest predictability results and their significance vary across sectors, meaning it becomes more informative in some sectors than others. There is no clear horizon that dominates statistical significance – each horizon has two significant sectors. Healthcare and information technology are significant among short-term horizons. These industries tend to have the most uncertainty and information asymmetry, supporting the idea that traders build their shorting strategies around short-term information, such as earnings weakness, analyst downgrades, negative news and others, whose effects are amplified in volatile industries. IT exhibits significant results because uncertainty is greater and fundamentals are not straightforward, enabling short sellers to more effectively recognize overvaluation. Furthermore, it is likely these two industries have greater representation in the 2000 firms, hence more observation points that are driving the significance of the results. Energy and consumer staples show positive coefficients at longer horizons. For energy the coefficient is large, in line with short sellers wrongly interpreting mean-reverting commodity-driven firms: they short during downturns but energy stocks bounce back with commodity cycles, mirroring a short squeeze. For consumer staples perhaps investors have wrong expectations about the industry by shorting defense stocks with stable cash flows, which reverses around 6-month horizon following the earnings cycles. The lack of significance in the remaining sectors may be driven by a lack of observations to justify the effect or these sectors are known to be more stable and predictive; hence, investors have more uniform expectations about the company's fundamentals.

Figure 5 shows the distribution of SIR by sector, which differs and the composition can influence the baseline long-short strategy: sectors which are naturally more shorted can dominate the long-short portfolio, becoming a strategy that longs or shorts disproportionately a single industry. The sector-neutral test addresses this concern through focusing on long-short strategies within specific industries, and finding positive and statistically significant results at all horizons (Figure 24). This implies that the short interest signal contains firm-level information within each sector, enabling traders to short weaker firms within a single sector. Sector-neutral implementation maintains the short-interest signal and reduces unintended industry risks during portfolio construction, improving the credibility of the strategy as a stock-selection signal.

4.3 Portfolio construction robustness

The baseline monthly long-short strategy produced returns and was significant up until the 6-month horizon; however, this baseline signal has to be manipulated to discover whether certain factors are driving the success behind the strategy (Figure 12). By changing sorting from deciles to quintiles we reduce concentration risk and any arising ranking noise through reducing cutoff sensitivity. The short-term significance of quintile sorting demonstrates the signal does not exist only in extreme deciles. Rebalancing the portfolio at a quarterly rate, which reduces turnover per rebalance, does not cause loss in the signal's significance. This finding is critical as it shows significant returns can be made through a more realistic strategy, which involves less frequent rebalancing. Furthermore, by excluding the 10th decile of the portfolio reveals if the strategy's viability is dependent on the most shorted stocks, also known as the stocks which have the highest borrowing fees and are most difficult to find; hence, there lies more predictive power in stocks which are difficult to obtain given the time and resources required to analyze these stocks and borrow the stock for shorting. Results return meaningful abnormal returns, even after excluding the 10th decile, suggesting the predictive power is not concentrated in this decile. Through these various tests we observe that changing the portfolio's structure can slightly influence the results, but it does not take away the significance of the positive abnormal results, demonstrating the strength of the shorting signal. Asquith et al. (2005) also finds strong results in the equally-weighted portfolios; however, recognizes that the most constrained stocks in these portfolio stocks are difficult to trade and their costs can undermine the initial profitability. To understand whether the returns are explained by the appropriate risk taken through the trading strategy, or if the strategy actually generates alpha, we analyze whether size, value, market beta or momentum factors explain the returns. After running CAPM, FF3 and the FF4 models we observe positive significant alphas, implying short interest reflects information other than those conveyed by the standard equity factors. The alphas decline with each model given new factors are added, especially momentum, but their significance remains: part of the strategy's performance may overlap with known factor exposures, but the short-interest signal retains explanatory power. These results are in line with what is found in literature, including Diether et al.'s (2008) paper, whereby abnormal returns are also not driven by FF3 and FF4 factors – the strategy conveys information that is particular to short selling.

4.4 Conditional sorts: where the signal is economically strongest

To better understand whether a specific firm characteristic drives the predictive power of short interest, stocks are sorted based on size, beta, liquidity and momentum and liquidity factors into quintiles to see which firm characteristics explain the returns (Figure 15).

Firstly, the SMB factor sort reveals that smallest stocks generate the largest annualized returns. This is likely driven by trading frictions and limits to arbitrage – there is less liquidity, search costs are higher, there is limited lending supply and less analyst coverage, making information incorporation a slower process. This is supported by Asquith et al. (2005) where the researchers find strong results amongst equal-weighted stocks, a method which gives greater weighting to smaller stocks. Schultz (2024) finds that stocks with higher fees are smaller, further explaining the difficulty of shorting smaller stocks. These restrictions may invite only the most sophisticated investors to short; however, even the highest restrictions can deter investors from exploiting the overvaluation, leading to a slow transition to the accurate value.

For momentum, the losing stock quintile exhibited the strongest results with highest annualized returns. This implies that short sellers are likely to identify companies which continue to decline rather than going against the most recent winners, which would imply a reversal – short sellers do not target companies necessarily with overreaction after these price rises.

Regarding beta, we see significant positive abnormal returns amongst the extremes of the betas – it only becomes insignificant for the middle quintile. The strongest effect still lies in the 5th quintile, with highest beta, indicating the signal is most informative in stocks with more valuation uncertainty, market sensitivity, and riskiness. The significance amongst other quintiles signifies that short interest does not only compensate for high systemic risk, as we see even the smallest beta stocks returning a significant positive abnormal return.

For liquidity we observe strongest results for least liquid stocks with great significance. This result coincides with the previous findings on the SMB, whereby less liquid stocks are also likely to be smaller-firm stocks, and predictive power increases amongst these firms as it becomes costly to borrow these stocks to short. However, the results are also significant and positive amongst other quintiles, with exception of the most liquidity stocks, which implies that the strategy is implementable as you can obtain abnormal returns on the strategy even for stocks that are relatively easy to short.

No single factor analyzed fully explains the types of stocks that experience better returns from the long-short strategy, indicating the strategy is implementable across different types of stocks in the market, but still with some limiting degree. Overall, the conditional sorts reveal that the short-interest signal is economically meaningful across different parts of the cross-section but is not uniform. The strongest effects are observed among small firms and past losers, suggesting that profitability is concentrated in firms with greater information asymmetry, financial distress, and stronger limits to arbitrage. In contrast, weaker results among large firms, past winners, and highly liquid stocks indicate that the signal is less informative where information is incorporated into prices more rapidly.

4.5 Transaction costs, borrow costs, and the gap between gross and net returns

Up to this point the regressions and long-short strategies have provided positive and significant abnormal returns, indicating a strong shorting signal; however, once borrowing costs and other restrictions are considered, the signal weakens, challenging the interpretation thus far. As shown in Figure 25, even with more moderate borrowing fee structures, the cost-adjusted results lose significance at a 10% level. The results align with what is found in literature: Jones and Lamont (2001) reveal that existing shorting frictions in the market enable overpricing to endure, while Saffi and Sigurdsson (2011) point out another important trading friction, whereby limited lending supply affects the efficiency of stock pricing. Overall, these findings reveal short interest as a predictive signal, but not as an effective tool for arbitrage exploitation. Raw returns are not sufficient to implement a profitable strategy, with borrowing fees, turnover and other restrictions affecting the results, enabling overvaluations to persist through limited market pricing correction.

4.6 Short-squeeze risk and filtered portfolios

Following the weakened long-short portfolio results after incorporating borrowing fees, filtering out firms with highest shorting can improve the strategy profitability. Stocks with high SI, high DTCR, and strong recent returns are removed from the portfolio as they can force short covering. Analyzing Figures 31 – 33, we see positive and significant results for the squeeze-filtered quintile monthly and quarterly strategies, with the former having greater significance. This result implies the construction of the short leg remains the greatest challenge while creating a long-short strategy due to the associated higher costs when shorting stocks. Through this method, the shorting signal is preserved while avoiding extreme adverse short-leg outcomes. Based on this outcome, portfolios which rely on shorting must be carefully constructed – short interest can be informative as an indicator for predictability, but it may not translate to a profitable strategy if borrowing fees overshadow the abnormal positive returns of the strategy.

The short-squeeze analysis reinforces that the short-interest signal is economically meaningful, but its practical implementation depends on the short leg's construction. Avoiding stocks with high SIR, high DTCR, and strong recent price increases improves net performance relative to the other transaction-cost-adjusted portfolios. However, even the squeeze-filtered strategy becomes statistically insignificant under more conservative cost assumptions, providing improved implementation and not a frictionless arbitrage opportunity.

4.7 Long-only implementation and portfolio exclusion

The transaction-cost results show that the direct long-short implementation becomes difficult once borrowing fees and short-side costs are included. To address this limitation, the paper also tests whether short interest can be used by a long-only investor through portfolio exclusion rather than shorting. This changes the interpretation of the signal: instead of using short interest to identify stocks to short, the strategy uses it to avoid weak long positions.

Figure 36 demonstrates that the portfolio which excludes high SIR and low momentum stocks produces the greatest returns, suggesting that SIR and other information can be used to construct a portfolio that beats the market. The highly shorted stocks and those that have low momentum have already been identified by investors as fundamentally weak, overvalued or likely to continue declining; hence, excluding these stocks in a long strategy reduces elements that would drag the alpha. Excluding top 20% SIR stocks consistently produces greater returns than excluding high SIR or low momentum stocks. The "or" classification may include more stocks than needed that could drag down the performance – low momentum may not effectively capture underperforming stocks, requiring short interest to also be considered to further support them as a bad position; hence, including stocks that are low momentum but not high SIR may exclude better stocks. The long-only benchmark earns an annualized net return of 18.85% with a Sharpe ratio of 0.79 (Figures 36 and 37). Excluding stocks that are both highly shorted and low momentum improves performance, generating an annualized net return of 19.54% and a Sharpe ratio of 0.83. The result is statistically significant, with a Newey-West t-statistic of 3.84. This strategy also slightly improves maximum drawdown relative to the equal-weighted benchmark.

However, it is also important to consider the performance of the strategy relative to the risk taken. As observed, in Figure 37, the strategy excluding high SIR and low momentum stocks produces the highest sharpe ratio in the 1- and 3-month horizon, making this more a short term strategy, while excluding SIR or low momentum becomes a long-term strategy. This may occur as low momentum becomes more informative in the long run and high SIR is not needed to reinforce the exclusion of bad stocks.

Considering long-short portfolio performance against long-only, net returns are higher with greater statistical significance. Even for the short-squeeze strategies where we saw positive net returns, the sensitivity to transaction costs complicates the ability to conclude on these strategies' profitability. Long-only strategy provides more certainty when building the strategy given the available data points.

5 Limitations

Despite the important findings uncovered and conclusions discussed above, the research paper encountered some limitations that must be addressed. Firstly, no direct stock-level borrowing-fee data was applied throughout the paper given the limited availability of such information within our setup. Two types of borrow-fee assumptions are considered, the main base case and the stress calibration which is more conservative. In both cases, post-fee portfolio significance and performance declined, as a cause determining the current borrowing fees becomes critical to evaluate whether the long-short portfolio is an actual implementable strategy or if it is just a signal. Furthermore, these assumptions may be too simple as they are likely to vary over time and by firm. The 2,000 universe makes the portfolio more investable but excludes microcaps, which are likely to have higher returns following a short – raw predictive power of short interest may be understated. By using the last trading day in December, we include firms that are bankrupt next year or have disappeared from the stock market. Though this reflects a likely universe the following year, it also reduces the availability of data as there will be fewer than 2,000 firms. One of the greatest limitations is the monthly frequency of short interest data. These short signals change daily, which may explain the weaker results in the change in SIR and DTCR as the variables are smoothed over the month. This also may not reflect the availability of data traders possess. With more sophisticated systems, investors are likely to have more frequent short interest information, which can be more informative than our current analysis. We relied on equal-weighted portfolios as this reduced rebalancing risk; however, it may be difficult to scale such portfolios due to the smaller companies: they have lower liquidity, can be difficult to borrow, alongside other trading restrictions. Also as noted in literature, value-weighted portfolios produce a less potent shorting signal, which can influence the outcome of our analysis. Our universe is US specific, which does provide extensive data for analysis, but the strategy may not be necessarily implementable globally due to differences in regulation, shorting restrictions and overall differences in the markets. The chosen time period also includes critical events that are likely to influence the results, and the smoothed-out results over averages can hide the signal's strength or weakness. Rapach et al.'s (2016) highlights these concerns, whereby return predictability can be limited to periods of macroeconomic distress. Lastly, we have the risk of omitted variables. Despite the inclusion of various time-, company-fixed effects, as well as other controls, not all variables were considered, meaning there could be other factors that still effectively explain the returns found in the regressions and strategies.

6 Conclusion

The thesis aimed to explore whether short interest can predict stock future returns, followed by if this signal can be used to construct a profitable and implementable portfolio, through the long-short strategy. A dynamic 2,000 company universe of the largest US company was used from 2007 to 2024 for OLS and panel regressions to establish the relationship, and further trading strategy construction to test for the robustness of the strategy, other control variables and most importantly the implementation of the strategy.

Following the model, we found SIR to be the strongest short-interest predictor, specifically that firms with high SIR underperformed in the future. Long-short portfolio of buying low-SIR stocks and shorting high-SIR stocks produced positive and significant gross returns, with most significant result at short- and medium-horizons. Factor-adjusted alphas remained positive and significant, and sectors did not influence the outcome of the long-short strategy. Despite the strong signals across the different robustness tests, the borrowing fee restrictions reduce the returns and their significance – net returns turned negative. These findings imply that short interest is informative and provides a strong signal; however, it is difficult to build a strategy that exploits this arbitrage. The squeeze-filtered strategies provided the most promising results by avoiding stocks that are most shorted, as they are associated with higher transaction costs that could reduce the returns on the short leg. This strategy produced positive and significant returns while still maintaining exposure to the short strategy. However, the short interest signal could also be used to exclude these stocks in a long-only strategy, where we also observed positive returns that beat the benchmark equity index.

The thesis explored a variety of regressions and portfolios to uncover the conclusion; however, there are some limitations to consider which may influence the result. Due to data limitations, stock-level borrow-fee data is not accessible, short interest data is only available bi-weekly, and the universe is focused on US companies whose data is available. Furthermore, omitted variables and the chosen time period can also limit the reliability of provided conclusions. For future research, these findings could be expanded internationally to gain a more global perspective, apply more frequent data to achieve more meaningful and significant results, use more variants of short interest for regressions and obtain more accurate transaction costs for the strategies.

The findings imply that short interest is an important signal for future return prediction, but not a simple arbitrage opportunity. Short interest should not be taken at face value to construct an implementable portfolio, but using various filters and restrictions we are able to construct a portfolio that still includes exposure to the short leg and can produce abnormal returns under certain assumptions, or can focus on the long leg while exploiting information about short interest.

References

Akbas, F., Boehmer, E., Erturk, B. and Sorescu, S. (2017). Short Interest, Returns, and Unfavorable Fundamental Information. *Financial Management*, 46(2), pp.455–486. doi:https://doi.org/10.1111/fima.12144.

Asquith, P., Pathak, P.A. and Ritter, J.R. (2005). Short interest, institutional ownership, and stock returns. *Journal of Financial Economics*, 78(2), pp.243–276. doi:https://doi.org/10.1016/j.jfineco.2005.01.001.

Bekjarovski, F. (2018). How do short selling costs and restrictions affect the profitability of stock anomalies? Tilburg University working paper.

Boehmer, E., Huszár, Z.R., Wang, Y., Zhang, X. and Zhang, X. (2021). Can Shorts Predict Returns? A Global Perspective. *The Review of Financial Studies*, 35(5), pp.2428–2463. doi:https://doi.org/10.1093/rfs/hhab079.

Branch, B. (1976). The Predictive Power of Stock Market Indicators. *The Journal of Financial and Quantitative Analysis*, 11(2). doi:https://doi.org/10.2307/2979054.

Compustat. Security Short Interest Database. S&P Global Market Intelligence / Wharton Research Data Services.

Center for Research in Security Prices (CRSP). Daily Stock File. Wharton Research Data Services.

Diether, K.B., Lee, K.-H. and Werner, I.M. (2008). Short-Sale Strategies and Return Predictability. *Review of Financial Studies*, 22(2), pp.575–607. doi:https://doi.org/10.1093/rfs/hhn047.

Duffie, D., Garleanu, N. and Pedersen, L.H. (2002). Securities Lending, Shorting, and Pricing. *Journal of Financial Economics*, 66(2-3). doi:https://doi.org/10.2139/ssrn.302015.

Kenneth R. French Data Library. Fama–French Factors and Momentum Factor. Dartmouth College.

Gorbenko, A. (2023). Short Interest and Aggregate Stock Returns: International Evidence. *The Review of Asset Pricing Studies*, 13, pp.691–733. doi:https://doi.org/10.1093/rapstu/raad007.

Grullon, G., Michenaud, S. and Weston, J.P. (2015). The Real Effects of Short-Selling Constraints. *The Review of Financial Studies*, 28(6), pp.1737–1767. doi:https://doi.org/10.1093/rfs/hhv013.

Hanna, M. (1976). A Stock Price Predictive Model Based on Changes in Ratios of Short Interest to Trading Volume. *The Journal of Financial and Quantitative Analysis*, 11(5). doi:https:

//doi.org/10.2307/2330585.

Horstmeyer, D., Wei, T. and Xia, J. (2023). Short Squeezes: A Four-Factor Model. [online] CFA Institute Research and Policy Center. Available at: https://rpc.cfainstitute.org/blogs/enterprising-investor/2023/short-squeezes-a-four-factor-model [Accessed 1 Jun. 2026].

Ji, C. (2024). Hedge funds are shorting these 7 stocks amid global market sell-off. [online] Business Insider. Available at: https://www.businessinsider.com/most-shorted-stocks-hedge-funds-global-market-sell-off-bofa-2024-8 [Accessed 1 Jun. 2026].

Jones, C.M. and Lamont, O.A. (2001). Short Sale Constraints And Stock Returns. *SSRN Electronic Journal*, NBER Working Paper No. w8494. doi:https://doi.org/10.2139/ssrn.281514.

Kelley, E.K. and Tetlock, P.C. (2017). Retail Short Selling and Stock Prices. *The Review of Financial Studies*, 30(3), pp.801–834. doi:https://doi.org/10.1093/rfs/hhw089.

Muravyev, D., Pearson, N.D. and Pollet, J.M. (2017). Understanding Returns to Short Selling Using Option-Implied Stock Borrowing Fees. Working paper.

Plender, J. (2021). GameStop is just latest sorry case of misallocated capital. [online] Financial Times. Available at: https://www.ft.com/content/ba875bce-0c7f-47f8-8c3f-5b98bd0b0b9b?syn-25a6b1a6=1 [Accessed 1 Jun. 2026].

Pollard, A. and Mourselas, C. (2026). Tech rally hands hedge funds biggest gains since 2020. [online] Financial Times. Available at: https://www.ft.com/content/df76569d-a840-42c4-8580-2d98bf1a16ce?syn-25a6b1a6=1 [Accessed 1 Jun. 2026].

Rapach, D., Ringgenberg, M. and Zhou, G. (2016). Short Interest and Aggregate Stock Returns. *Journal of Financial Economics*, 121, pp.46–65. doi:https://doi.org/10.2139/ssrn.2675298.

Saffi, P.A.C. and Sigurdsson, K. (2011). Price Efficiency and Short Selling. *The Review of Financial Studies*, 24(3).

Schultz, P. (2024). Short Sale Constraints and Stock Returns: 1926–2023. University of Notre Dame working paper.

Zhu, Z., DUAN, X. and Tu, J. (2019). The trend in short selling and the cross section of stock returns. *Annals of Economics and Finance*, 20(2), pp.565–586.

A Appendix Figures

A.1 Baseline Predictive Regressions

Figure 1: Pooled Multivariate OLS Regressions

	1M	3M	6M	12M
SIR	-0.0525**	-0.2217***	-0.3422*	-0.2924
	(0.0168)	(0.0040)	(0.0665)	(0.4068)
ΔSIR	0.1488*	0.2906	0.5436	-0.7730
	(0.0817)	(0.3100)	(0.4340)	(0.3770)
DTCR	-0.0003	-0.0004	-0.0014	-0.0052**
	(0.1060)	(0.3950)	(0.1770)	(0.0310)
Size	-0.0014	-0.0042	-0.0167	-0.0502
	(0.2780)	(0.3640)	(0.1960)	(0.1120)
Turnover	-0.0786	0.1851	-0.0952	0.1350
	(0.6650)	(0.6320)	(0.9190)	(0.9550)
Average daily volume	2.069×10^{-11}	8.768×10^{-11}	2.231×10^{-9}	9.021×10^{-9}
	(0.9410)	(0.9340)	(0.4420)	(0.1700)
Observations	433,412	142,887	69,505	32,817
R^2	0.0013	0.0036	0.0072	0.0183
Month-clustered SE	Yes	Yes	Yes	Yes
Non-overlapping sample	No	Yes	Yes	Yes

Notes: The dependent variable is the winsorized future stock return over each horizon.

P-values are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

Figure 2: Two-Way Fixed-Effects Panel Regressions

	1M	3M	6M	12M
SIR	-0.0549***	-0.2261***	-0.3179***	-0.5710**
	(0.0017)	($<$ 0.0001)	(0.0022)	(0.0403)
ΔSIR	0.0597***	0.0587	-0.0428	0.2193
	(0.0065)	(0.3931)	(0.6935)	(0.3181)
DTCR	-8.720×10^{-5}	0.0002	-0.0002	-0.0005
	(0.4889)	(0.6534)	(0.7888)	(0.7612)
Size	-0.0195***	-0.0576***	-0.1240***	-0.2662***
	($<$ 0.0001)	($<$ 0.0001)	($<$ 0.0001)	($<$ 0.0001)
Turnover	-0.1302	0.0273	-0.6043	-1.0841
	(0.1762)	(0.9201)	(0.3360)	(0.4540)
Average daily volume	-1.113×10^{-10}	-2.622×10^{-10}	1.858×10^{-10}	1.876×10^{-9}
	(0.6514)	(0.7114)	(0.8906)	(0.4855)
Observations	433,412	142,887	69,505	32,817
Entities	4,711	4,676	4,621	4,032
Time periods	295	235	213	164
R^2	0.0082	0.0236	0.0520	0.1117
Firm fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Two-way clustered SE	Yes	Yes	Yes	Yes
Non-overlapping sample	No	Yes	Yes	Yes

Notes: The dependent variable is the winsorized future stock return over each horizon.

P-values are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

Figure 3: Univariate Two-Way Fixed-Effects Panel Regressions

Variable	Statistic	1M	3M	6M	12M
SIR	Coefficient	-0.0116	-0.0382	-0.0220	0.0614
	P-value	0.4088	0.3716	0.7717	0.7612
ΔSIR	Coefficient	0.0260	-0.0512	-0.1813	0.0473
	P-value	0.2423	0.4410	0.1032	0.7594
DTCR	Coefficient	2.000×10^{-5}	-5.000×10^{-6}	6.260×10^{-4}	0.0030***
	P-value	0.8307	0.9853	0.1599	0.0076
Size	Coefficient	-0.0183***	-0.0550***	-0.1192***	-0.2582***
	P-value	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Turnover	Coefficient	-0.0913	-0.3291	-0.8306	-2.2480**
	P-value	0.2750	0.1801	0.2116	0.0361
Average daily volume	Coefficient	≈ 0	≈ 0*	≈ 0*	≈ 0
	P-value	0.1320	0.0716	0.0946	0.1224
Observations for SIR		438,146	143,482	69,804	32,952
Observations for ΔSIR		433,412	142,887	69,505	32,817
Firm fixed effects		Yes	Yes	Yes	Yes
Time fixed effects		Yes	Yes	Yes	Yes

Notes: Each coefficient comes from a separate univariate panel regression with firm and time fixed effects.

Average daily volume is measured in raw share-volume units, so its coefficient is close to zero at the displayed scale.

P-values are reported below coefficients. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

A.2 Sector-Level Predictive Regressions

Figure 4: Sector-Level Univariate Panel Regressions: Coefficient on SIR

Sector	1M	3M	6M	12M
Communication Services	-0.0180	-0.0240	-0.0028	0.1392
Consumer Discretionary	-0.0149	-0.0503	-0.0035	0.1840
Consumer Staples	0.0227	0.0765	0.2807*	0.3791
Energy	0.0087	0.0085	0.2506	0.7506**
Financials	0.0132	0.0372	0.0493	0.2372
Health Care	-0.0352**	-0.0838*	-0.0760	-0.2491
Industrials	0.0178	0.0116	0.0197	0.1004
Information Technology	-0.0413***	-0.1193***	-0.2131**	-0.4022*
Materials	-0.0229	-0.0496	0.0081	0.3479
Real Estate	-0.0000	0.0484	0.2050	1.3402
Utilities	0.0109	0.1415	-0.0573	-0.1503

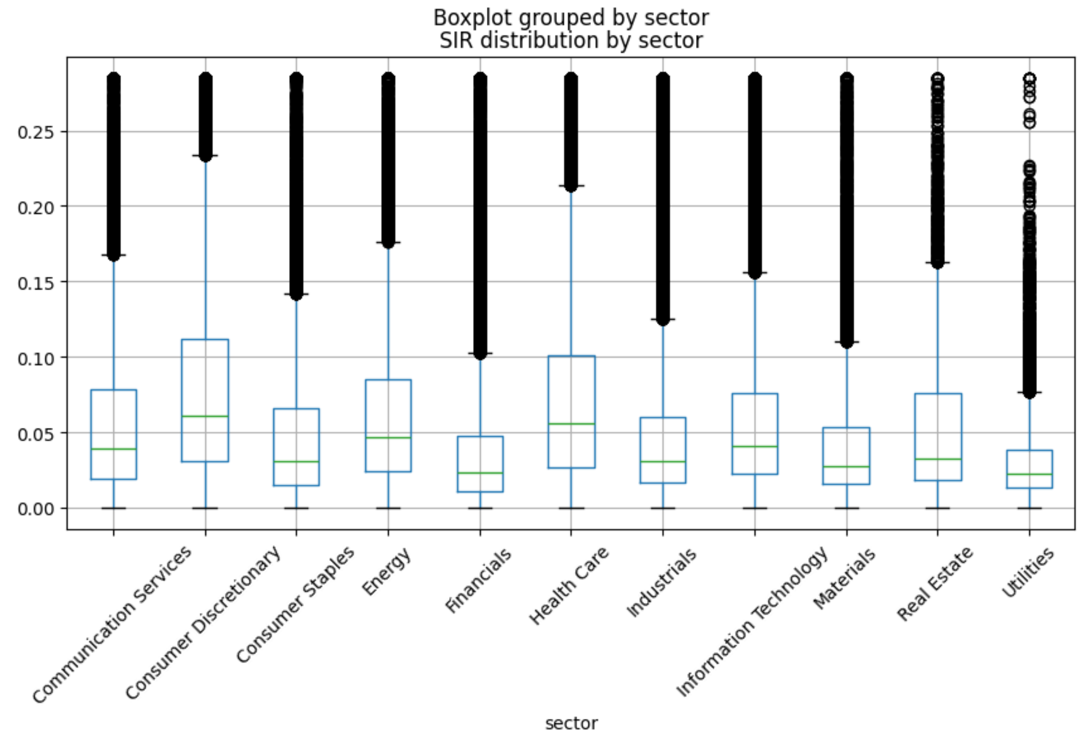
Notes: This table reports coefficients on SIR from sector-level univariate panel regressions.

Each regression is estimated separately within a GICS sector and includes firm and time fixed effects.

Standard errors are clustered at the firm level.

*, **, and *** indicate significance at the 10%, 5%, and 1% levels.

Figure 5: Distribution of Short Interest Ratio by Sector



A.3 Baseline Long-Short Strategy

Figure 6: Baseline Long-Short Strategy: Low SIR Minus High SIR

Horizon	N	Mean LS Return	Newey-West t-stat	Newey-West p-value
1M	227	0.0073	3.044	0.0023
3M	225	0.0185	2.807	0.0050
6M	222	0.0270	1.965	0.0494
12M	216	0.0280	1.064	0.2876

Notes: This table reports the holding period return of a strategy long the lowest-SIR decile and short the highest-SIR decile. Returns are equal-weighted within deciles. Newey–West standard errors use lags equal to the return horizon.

Figure 7: Baseline 1 month Long-Short Strategy versus Market Benchmark

Metric	Long-Short Strategy	Market Benchmark
Mean monthly return	0.0073	0.0094
Annualized return	0.0917	0.1192
Monthly volatility	0.0357	0.0593
Annualized volatility	0.1237	0.2053
Sharpe ratio	0.7117	0.5511
Maximum drawdown	-0.3963	-0.5545
Correlation with market	-0.0058	1.0000

Notes: The market benchmark is the equal-weighted average return of stocks in the monthly portfolio sample.

Figure 8: Cumulative Return of the Long Low-SIR, Short High-SIR Strategy: 1 month Horizon

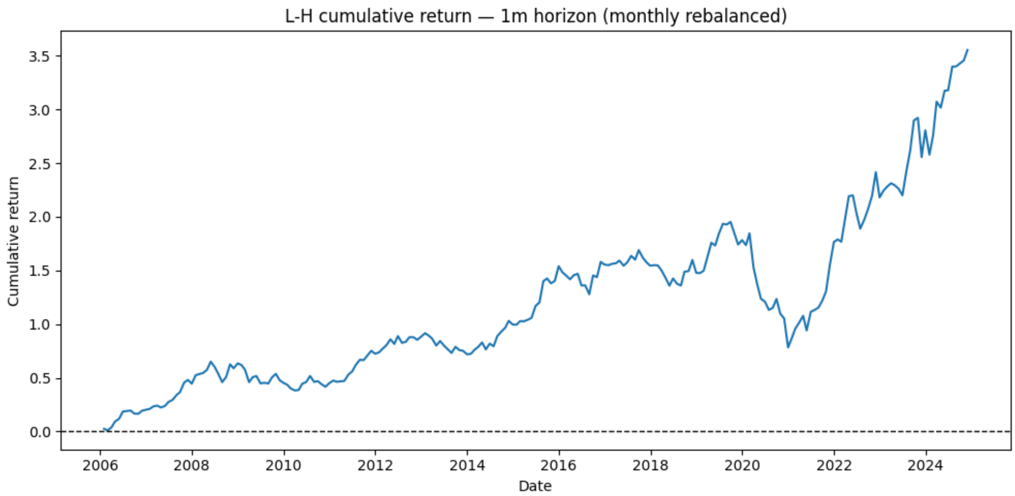


Figure 9: Cumulative Return of the Long Low-SIR, Short High-SIR Strategy: 3 month Horizon

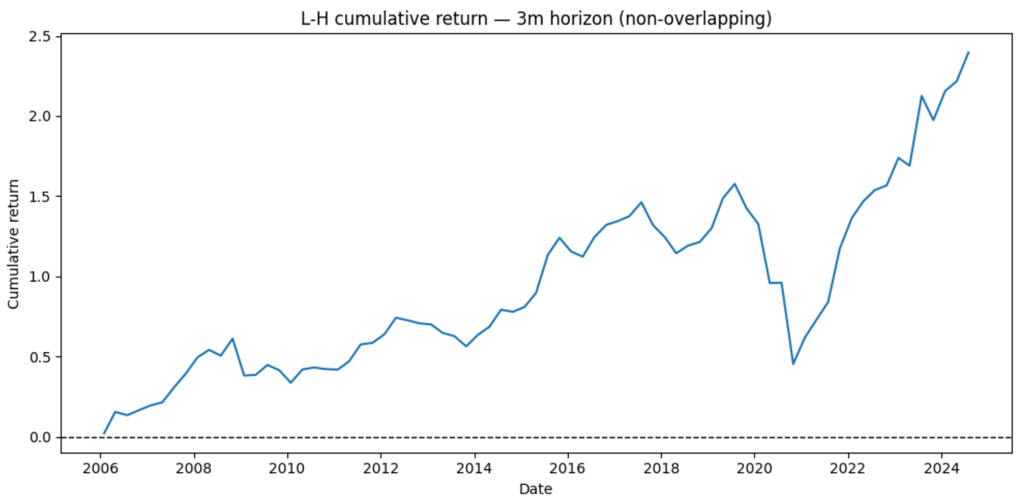


Figure 10: Cumulative Return of the Long Low-SIR, Short High-SIR Strategy: 6 month Horizon

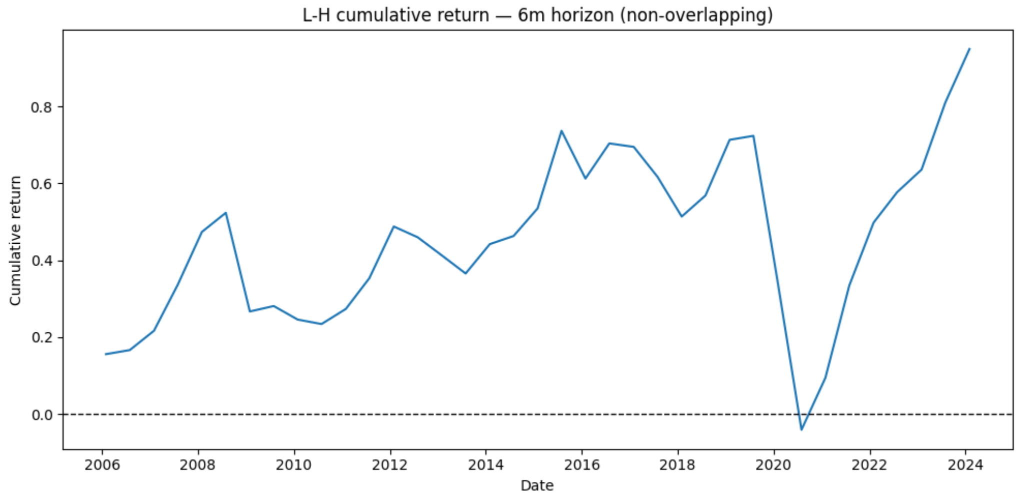
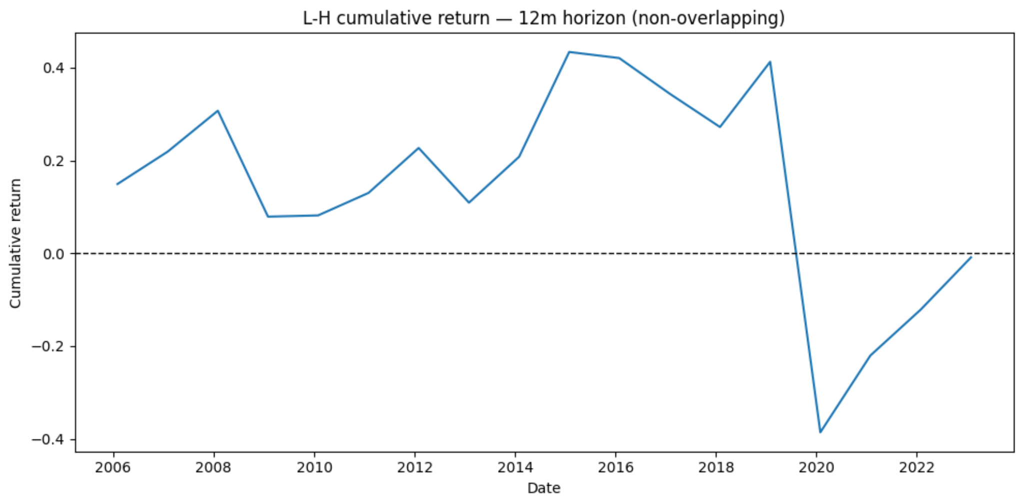


Figure 11: Cumulative Return of the Long Low-SIR, Short High-SIR Strategy: 12 month Horizon



A.4 Implementation Robustness and Transaction Costs

Figure 12: Implementation Robustness: Annualized Gross Returns and Turnover

Horizon	Strategy	Annualized Return	NW t-stat	NW p-value	Turnover per Rebalance	Monthly-Equivalent Turnover
1M	Baseline monthly decile	0.0917	3.044	0.0023	0.4194	0.4194
1M	Baseline quarterly decile	0.1313	2.325	0.0201	0.6256	0.2085
1M	Monthly quintile	0.0768	2.900	0.0037	0.2998	0.2998
1M	Quarterly quintile	0.1209	2.602	0.0093	0.4868	0.1623
1M	Exclude D10 monthly quintile	0.0446	2.214	0.0268	0.4025	0.4025
1M	Exclude D10 quarterly quintile	0.0818	2.571	0.0102	0.6476	0.2159
3M	Baseline monthly decile	0.0763	2.807	0.0050	0.4193	0.4193
3M	Baseline quarterly decile	0.0844	2.619	0.0088	0.6280	0.2093
3M	Monthly quintile	0.0642	2.780	0.0054	0.2992	0.2992
3M	Quarterly quintile	0.0663	2.368	0.0179	0.4906	0.1635
3M	Exclude D10 monthly quintile	0.0344	1.958	0.0502	0.4024	0.4024
3M	Exclude D10 quarterly quintile	0.0326	1.496	0.1348	0.6471	0.2157
6M	Baseline monthly decile	0.0548	1.965	0.0494	0.4204	0.4204
6M	Baseline quarterly decile	0.0401	1.364	0.1727	0.6253	0.2084
6M	Monthly quintile	0.0527	2.247	0.0246	0.2991	0.2991
6M	Quarterly quintile	0.0450	1.787	0.0740	0.4871	0.1624
6M	Exclude D10 monthly quintile	0.0326	1.813	0.0698	0.4025	0.4025
6M	Exclude D10 quarterly quintile	0.0329	1.724	0.0847	0.6425	0.2142
12M	Baseline monthly decile	0.0280	1.064	0.2876	0.4155	0.4155
12M	Baseline quarterly decile	0.0251	0.960	0.3370	0.6101	0.2034
12M	Monthly quintile	0.0318	1.398	0.1621	0.2933	0.2933
12M	Quarterly quintile	0.0300	1.250	0.2113	0.4731	0.1577
12M	Exclude D10 monthly quintile	0.0234	1.345	0.1786	0.3983	0.3983
12M	Exclude D10 quarterly quintile	0.0230	1.214	0.2248	0.6302	0.2101

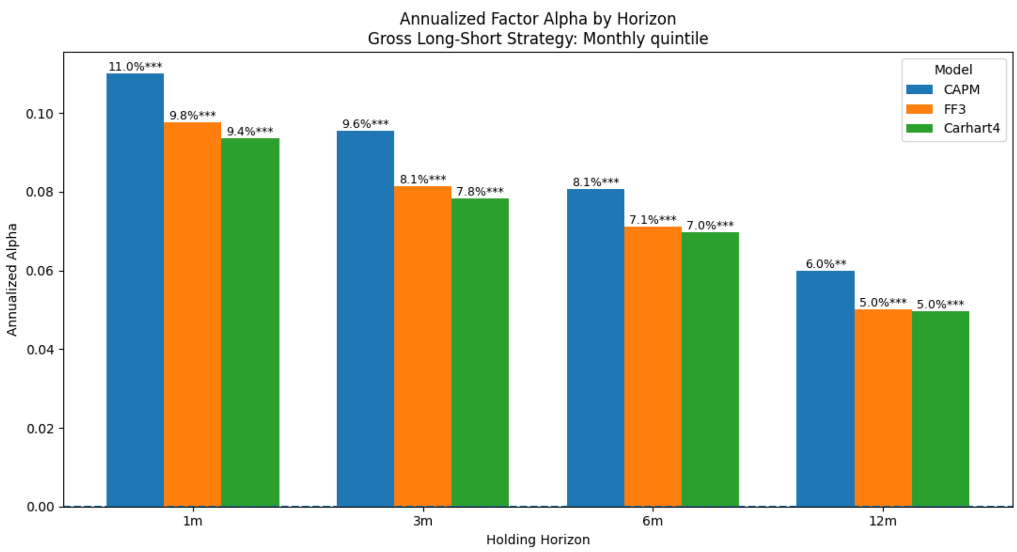
Notes: This table reports annualized gross long-short returns before explicit transaction costs. The underlying mean returns are holding-period returns and are annualized as $(1 + \bar{R}_h)^{12/h} - 1$, where h is the holding horizon in months.

Figure 13: Factor-Adjusted Performance: Monthly Quintile Strategy

Horizon	Model	Alpha	Annualized Alpha	Alpha t-stat	Alpha p-value	R^2
1M	CAPM	0.0087	0.1100	4.409	0.0000	0.2057
1M	FF3	0.0078	0.0976	4.848	0.0000	0.5303
1M	Carhart4	0.0075	0.0936	4.720	0.0000	0.5583
3M	CAPM	0.0231	0.0955	4.067	0.0000	0.1756
3M	FF3	0.0197	0.0813	5.124	0.0000	0.5670
3M	Carhart4	0.0190	0.0783	5.007	0.0000	0.5808
6M	CAPM	0.0395	0.0807	3.254	0.0011	0.1612
6M	FF3	0.0349	0.0711	6.335	0.0000	0.6507
6M	Carhart4	0.0342	0.0696	6.602	0.0000	0.6540
12M	CAPM	0.0598	0.0598	2.212	0.0270	0.1597
12M	FF3	0.0502	0.0502	4.769	0.0000	0.7102
12M	Carhart4	0.0497	0.0497	4.996	0.0000	0.7109

Notes: This table reports factor-adjusted alphas for the gross monthly quintile long-short strategy before transaction costs.

Figure 14: Annualized Factor Alpha by Holding Horizon



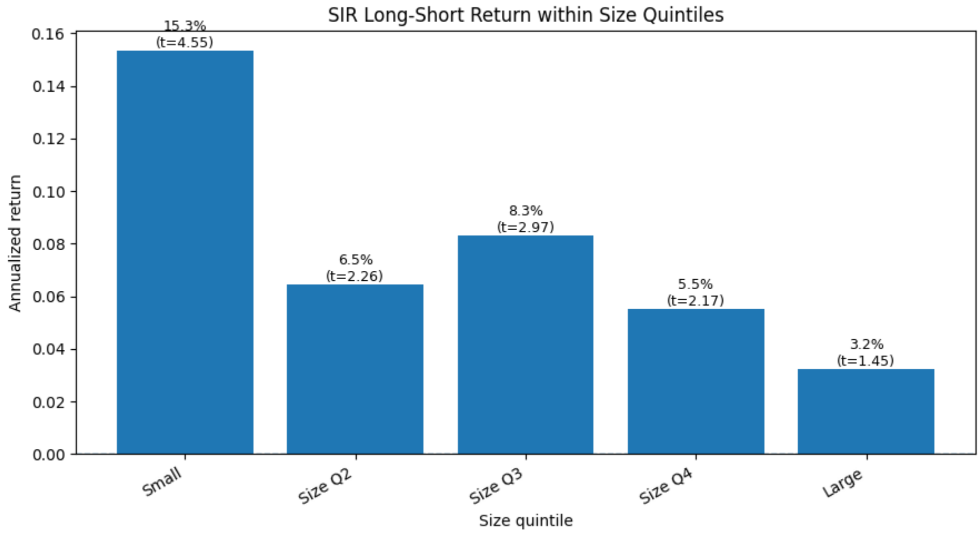
Notes: This figure reports annualized factor alphas for the gross monthly quintile long-short strategy before transaction costs. The strategy goes long the lowest-SIR quintile and short the highest-SIR quintile. Alphas are estimated using CAPM, Fama–French three-factor, and Carhart four-factor models. Significance stars are based on Newey–West p-values.

Figure 15: Conditional SIR Sorts by Firm Characteristics

Characteristic	Group	Mean Monthly Return	Annualized Return	NW t-stat	NW p-value
Size	Small	0.0120	0.1532	4.546	0.0000
Size	Size Q2	0.0052	0.0645	2.263	0.0236
Size	Size Q3	0.0067	0.0832	2.974	0.0029
Size	Size Q4	0.0045	0.0552	2.169	0.0301
Size	Large	0.0027	0.0323	1.451	0.1468
Momentum	Losers	0.0114	0.1462	4.284	0.0000
Momentum	Momentum Q2	0.0043	0.0523	2.028	0.0425
Momentum	Momentum Q3	0.0016	0.0199	0.930	0.3523
Momentum	Momentum Q4	0.0010	0.0124	0.557	0.5775
Momentum	Winners	0.0011	0.0136	0.513	0.6081
Market beta	Low beta	0.0053	0.0653	2.176	0.0295
Market beta	Beta Q2	0.0043	0.0534	2.106	0.0352
Market beta	Beta Q3	0.0024	0.0294	1.222	0.2218
Market beta	Beta Q4	0.0044	0.0545	2.076	0.0379
Market beta	High beta	0.0074	0.0923	2.679	0.0074
Liquidity	Low liquidity	0.0068	0.0845	3.156	0.0016
Liquidity	Liquidity Q2	0.0090	0.1130	3.767	0.0002
Liquidity	Liquidity Q3	0.0076	0.0948	3.224	0.0013
Liquidity	Liquidity Q4	0.0063	0.0778	2.758	0.0058
Liquidity	High liquidity	0.0043	0.0535	1.584	0.1133

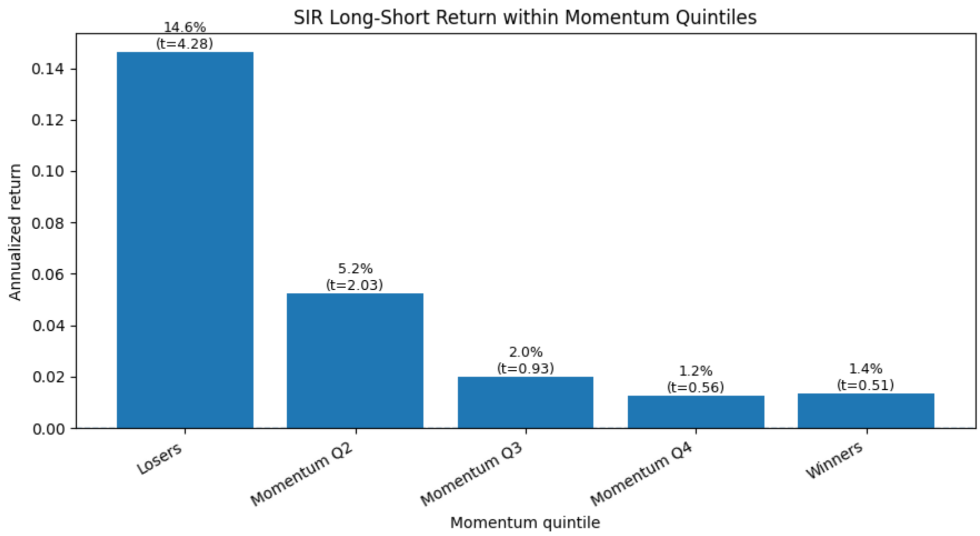
Notes: This table reports conditional SIR portfolio sorts by firm characteristics. At each month, stocks are first sorted into quintiles based on the characteristic shown in the first column. Within each characteristic quintile, stocks are then sorted into SIR quintiles.

Figure 16: SIR Long-Short Returns within Size Quintiles



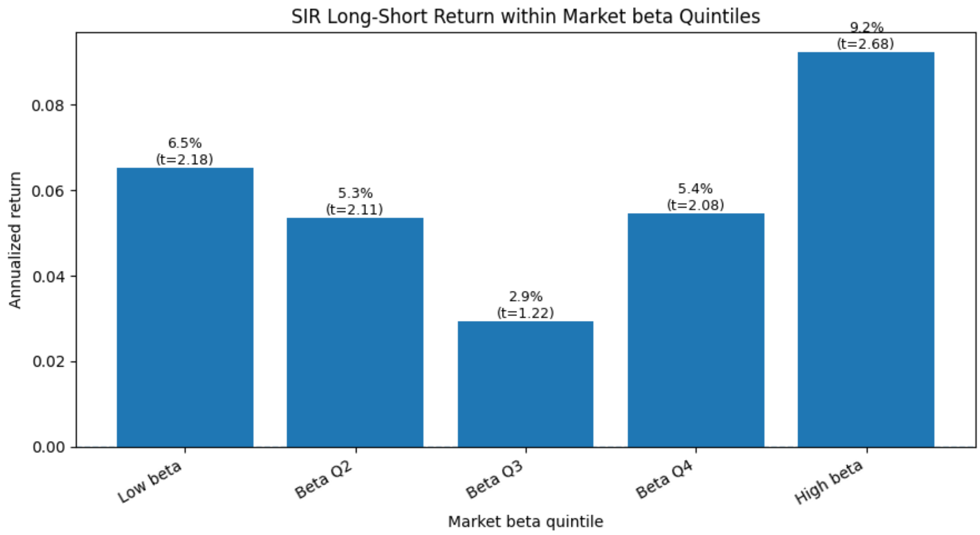
Notes: This figure reports annualized gross 1-month long-short returns within size quintiles. The strategy goes long the lowest-SIR quintile and short the highest-SIR quintile within each size group. Returns are equal-weighted and before transaction costs. Newey–West t-statistics are shown in parentheses.

Figure 17: SIR Long-Short Returns within Momentum Quintiles



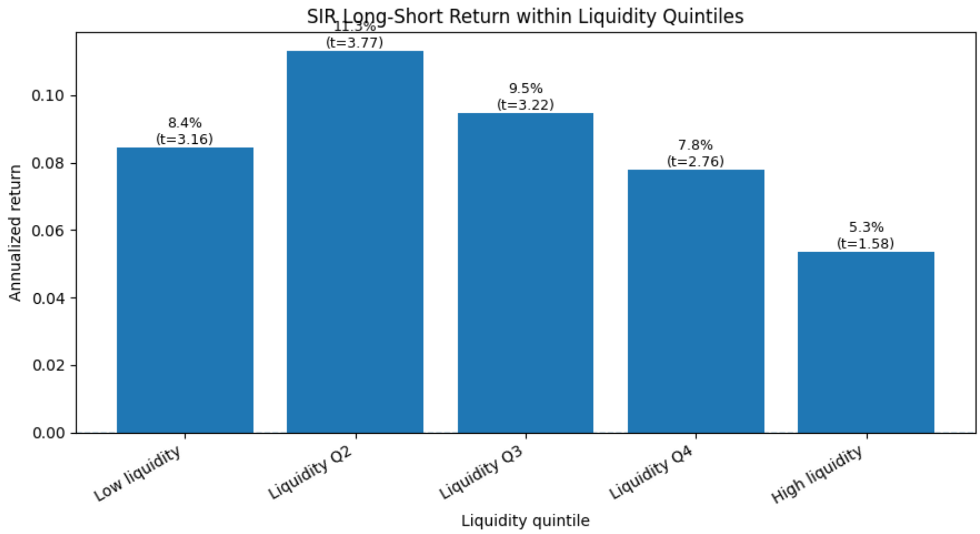
Notes: This figure reports annualized gross 1-month long-short returns within momentum quintiles. The strategy goes long the lowest-SIR quintile and short the highest-SIR quintile within each momentum group. Returns are equal-weighted and before transaction costs. Newey–West t-statistics are shown in parentheses.

Figure 18: SIR Long-Short Returns within Market Beta Quintiles



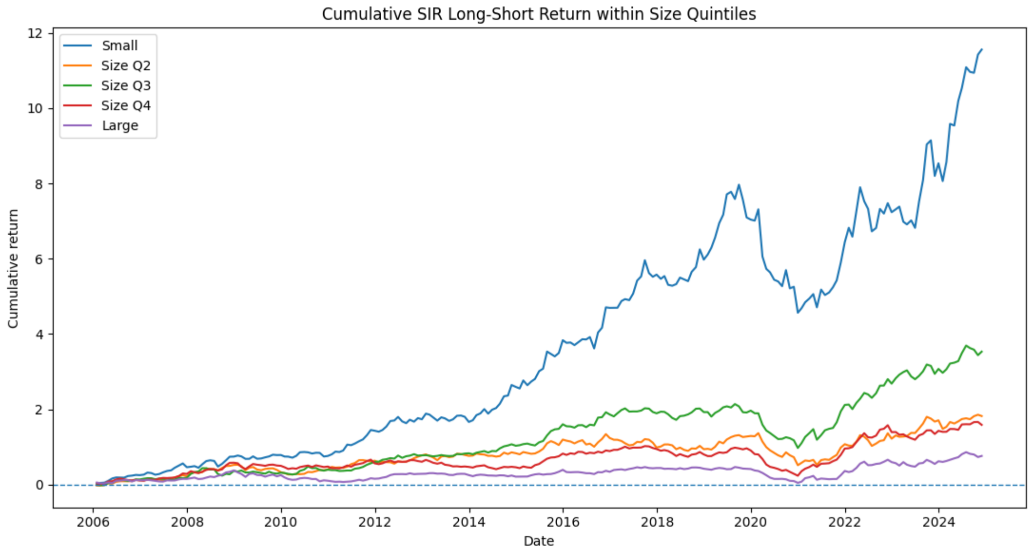
Notes: This figure reports annualized gross 1-month long-short returns within market beta quintiles. The strategy goes long the lowest-SIR quintile and short the highest-SIR quintile within each beta group. Returns are equal-weighted and before transaction costs. Newey–West t-statistics are shown in parentheses.

Figure 19: SIR Long-Short Returns within Liquidity Quintiles



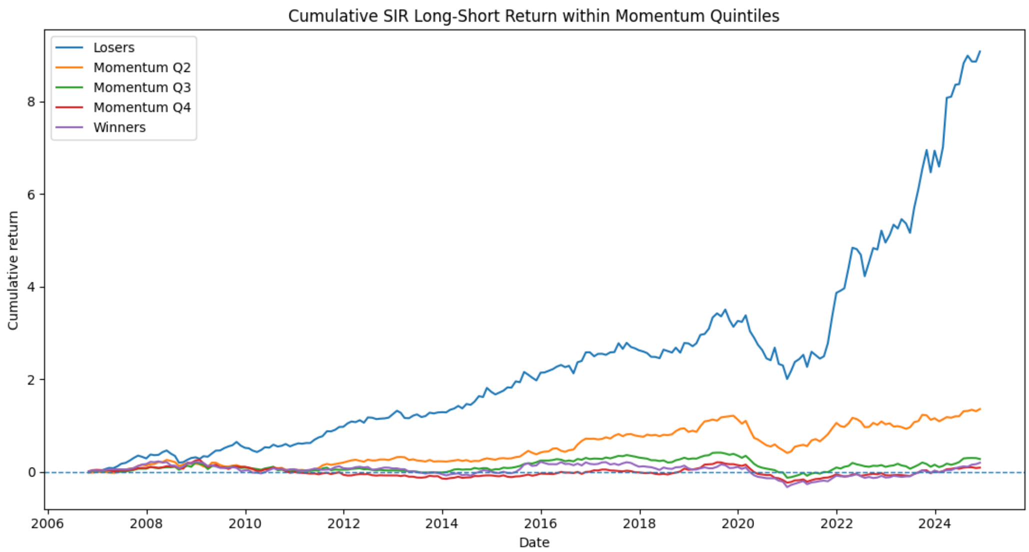
Notes: This figure reports annualized gross 1-month long-short returns within liquidity quintiles. The strategy goes long the lowest-SIR quintile and short the highest-SIR quintile within each liquidity group. Returns are equal-weighted and before transaction costs. Newey–West t-statistics are shown in parentheses.

Figure 20: Cumulative SIR Long-Short Returns within Size Quintiles



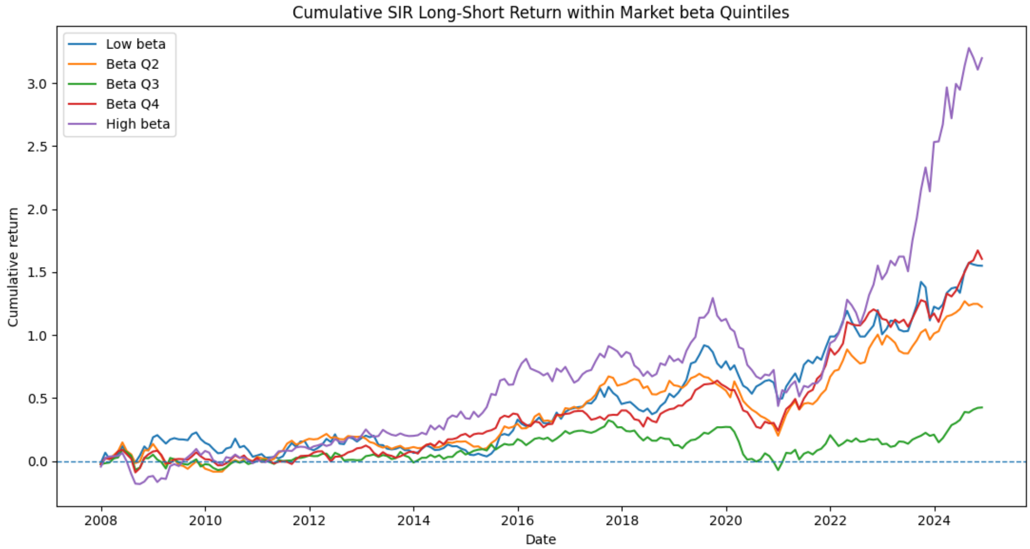
Notes: This figure plots cumulative gross 1-month long-short returns within size quintiles. The strategy goes long the lowest-SIR quintile and short the highest-SIR quintile within each size group. Returns are equal-weighted and before transaction costs.

Figure 21: Cumulative SIR Long-Short Returns within Momentum Quintiles



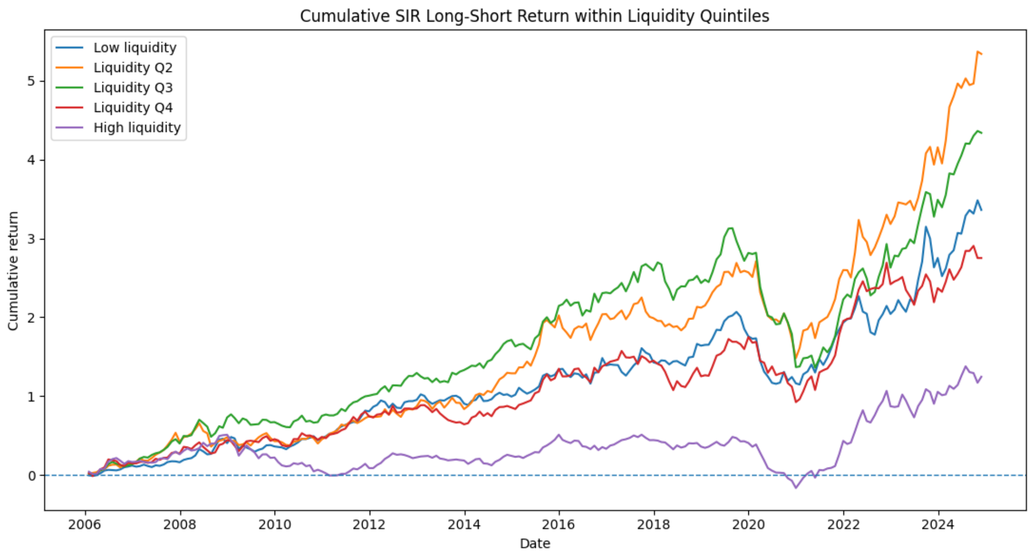
Notes: This figure plots cumulative gross 1-month long-short returns within momentum quintiles. The strategy goes long the lowest-SIR quintile and short the highest-SIR quintile within each momentum group. Returns are equal-weighted and before transaction costs.

Figure 22: Cumulative SIR Long-Short Returns within Market Beta Quintiles



Notes: This figure plots cumulative gross 1-month long-short returns within market beta quintiles. The strategy goes long the lowest-SIR quintile and short the highest-SIR quintile within each beta group. Returns are equal-weighted and before transaction costs.

Figure 23: Cumulative SIR Long-Short Returns within Liquidity Quintiles



Notes: This figure plots cumulative gross 1-month long-short returns within liquidity quintiles. The strategy goes long the lowest-SIR quintile and short the highest-SIR quintile within each liquidity group. Returns are equal-weighted and before transaction costs.

Figure 24: Sector-Neutral SIR Long-Short Strategy

Horizon	Mean Return	Annualized Return	NW t-stat	NW p-value	N
1M	0.0056	0.0698	3.430	0.0006	227
3M	0.0150	0.0615	3.732	0.0002	225
6M	0.0248	0.0501	3.057	0.0022	222
12M	0.0315	0.0315	2.028	0.0425	216

Notes: This table reports gross sector-neutral long-short returns before transaction costs.

Figure 25: Cost-Adjusted Results under the Main Borrow-Cost Calibration

Horizon	Strategy	Borrow Fee	Gross Mean Return	Net Mean Return	Gross NW t-stat	Net NW t-stat	Net NW p-value	Avg. Trading Cost	Borrow Cost
1M	Baseline decile monthly	5%	0.0073	0.0018	3.044	0.738	0.4606	0.0014	0.0042
1M	Baseline decile quarterly	5%	0.0103	0.0040	2.325	0.886	0.3758	0.0022	0.0042
1M	Quintile monthly	3%	0.0062	0.0027	2.900	1.246	0.2129	0.0010	0.0025
1M	Quintile quarterly	3%	0.0096	0.0053	2.602	1.442	0.1492	0.0017	0.0025
1M	Exclude D10 quintile monthly	1%	0.0036	0.0013	2.214	0.792	0.4284	0.0015	0.0008
1M	Exclude D10 quintile quarterly	1%	0.0066	0.0032	2.571	1.260	0.2078	0.0025	0.0008
3M	Baseline decile monthly	5%	0.0185	0.0047	2.807	0.705	0.4811	0.0014	0.0125
3M	Baseline decile quarterly	5%	0.0205	0.0057	2.619	0.735	0.4621	0.0022	0.0125
3M	Quintile monthly	3%	0.0157	0.0072	2.780	1.270	0.2042	0.0010	0.0075
3M	Quintile quarterly	3%	0.0162	0.0069	2.368	1.014	0.3105	0.0018	0.0075
3M	Exclude D10 quintile monthly	1%	0.0085	0.0045	1.958	1.034	0.3009	0.0015	0.0025
3M	Exclude D10 quintile quarterly	1%	0.0081	0.0031	1.496	0.568	0.5704	0.0025	0.0025
6M	Baseline decile monthly	5%	0.0270	0.0006	1.965	0.045	0.9644	0.0014	0.0250
6M	Baseline decile quarterly	5%	0.0198	-0.0074	1.364	-0.506	0.6127	0.0022	0.0250
6M	Quintile monthly	3%	0.0260	0.0100	2.247	0.864	0.3878	0.0010	0.0150
6M	Quintile quarterly	3%	0.0223	0.0055	1.787	0.443	0.6575	0.0017	0.0150
6M	Exclude D10 quintile monthly	1%	0.0162	0.0097	1.813	1.084	0.2784	0.0015	0.0050
6M	Exclude D10 quintile quarterly	1%	0.0163	0.0088	1.724	0.934	0.3500	0.0025	0.0050
12M	Baseline decile monthly	5%	0.0280	-0.0234	1.064	-0.891	0.3729	0.0014	0.0500
12M	Baseline decile quarterly	5%	0.0251	-0.0270	0.960	-1.032	0.3020	0.0021	0.0500
12M	Quintile monthly	3%	0.0318	0.0008	1.398	0.036	0.9710	0.0010	0.0300
12M	Quintile quarterly	3%	0.0300	-0.0017	1.250	-0.069	0.9446	0.0017	0.0300
12M	Exclude D10 quintile monthly	1%	0.0234	0.0119	1.345	0.686	0.4930	0.0015	0.0100
12M	Exclude D10 quintile quarterly	1%	0.0230	0.0105	1.214	0.557	0.5774	0.0024	0.0100

Notes: This table reports cost-adjusted long-short returns under the main calibration.

Figure 26: Cost-Adjusted Results under the Stress Borrow-Cost Calibration

Horizon	Strategy	Borrow Fee	Net Mean Return	Net NW t-stat	Net NW p-value
1M	Baseline decile monthly	10%	-0.0024	-0.990	0.3221
1M	Baseline decile quarterly	10%	-0.0002	-0.048	0.9621
1M	Quintile monthly	5%	0.0010	0.465	0.6421
1M	Quintile quarterly	5%	0.0036	0.990	0.3223
1M	Exclude D10 quintile monthly	3%	-0.0004	-0.222	0.8247
1M	Exclude D10 quintile quarterly	3%	0.0016	0.611	0.5415
3M	Baseline decile monthly	10%	-0.0078	-1.187	0.2353
3M	Baseline decile quarterly	10%	-0.0068	-0.864	0.3877
3M	Quintile monthly	5%	0.0022	0.383	0.7015
3M	Quintile quarterly	5%	0.0019	0.282	0.7780
3M	Exclude D10 quintile monthly	3%	-0.0005	-0.119	0.9056
3M	Exclude D10 quintile quarterly	3%	-0.0019	-0.362	0.7176
6M	Baseline decile monthly	10%	-0.0244	-1.775	0.0759
6M	Baseline decile quarterly	10%	-0.0324	-2.226	0.0260
6M	Quintile monthly	5%	-0.0000	-0.000	0.9997
6M	Quintile quarterly	5%	-0.0045	-0.360	0.7186
6M	Exclude D10 quintile monthly	3%	-0.0003	-0.038	0.9698
6M	Exclude D10 quintile quarterly	3%	-0.0012	-0.122	0.9031
12M	Baseline decile monthly	10%	-0.0734	-2.793	0.0052
12M	Baseline decile quarterly	10%	-0.0770	-2.943	0.0033
12M	Quintile monthly	5%	-0.0192	-0.842	0.3996
12M	Quintile quarterly	5%	-0.0217	-0.903	0.3668
12M	Exclude D10 quintile monthly	3%	-0.0081	-0.463	0.6437
12M	Exclude D10 quintile quarterly	3%	-0.0095	-0.499	0.6177

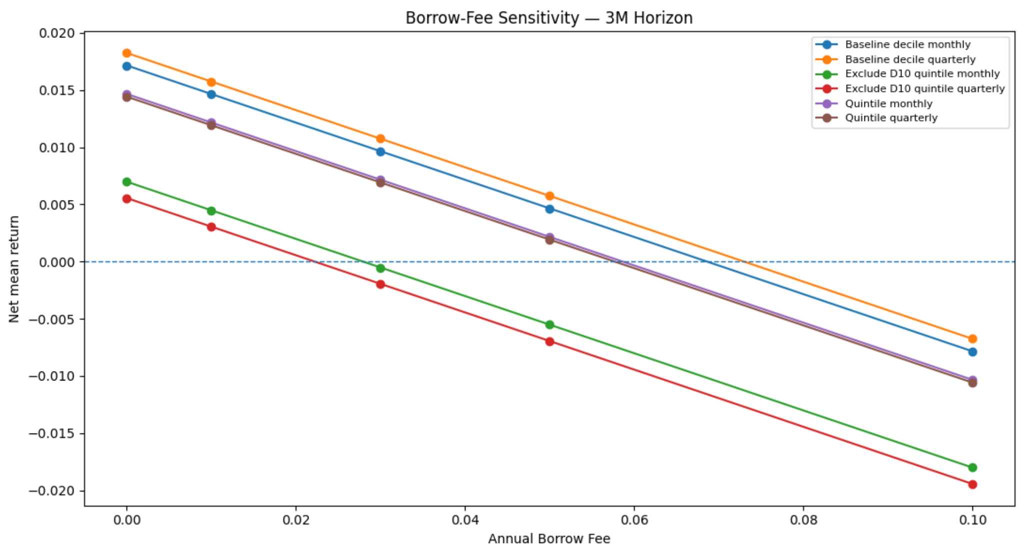
Notes: This table reports net returns under the stress calibration.

Figure 27: Borrow-Fee Sensitivity at the 3-Month Horizon

Strategy	0%	1%	3%	5%	10%
Panel A: Net Mean Return					
Baseline decile monthly	0.0172	0.0147	0.0097	0.0047	-0.0078
Baseline decile quarterly	0.0182	0.0157	0.0107	0.0057	-0.0068
Quintile monthly	0.0147	0.0122	0.0072	0.0022	-0.0103
Quintile quarterly	0.0144	0.0119	0.0069	0.0019	-0.0106
Exclude D10 quintile monthly	0.0070	0.0045	-0.0005	-0.0055	-0.0180
Exclude D10 quintile quarterly	0.0056	0.0031	-0.0019	-0.0069	-0.0194
Panel B: Newey–West t-statistic					
Baseline decile monthly	2.596	2.218	1.461	0.705	-1.187
Baseline decile quarterly	2.335	2.015	1.375	0.735	-0.864
Quintile monthly	2.599	2.156	1.270	0.383	-1.832
Quintile quarterly	2.113	1.746	1.014	0.282	-1.549
Exclude D10 quintile monthly	1.611	1.034	-0.119	-1.272	-4.154
Exclude D10 quintile quarterly	1.032	0.568	-0.362	-1.291	-3.614

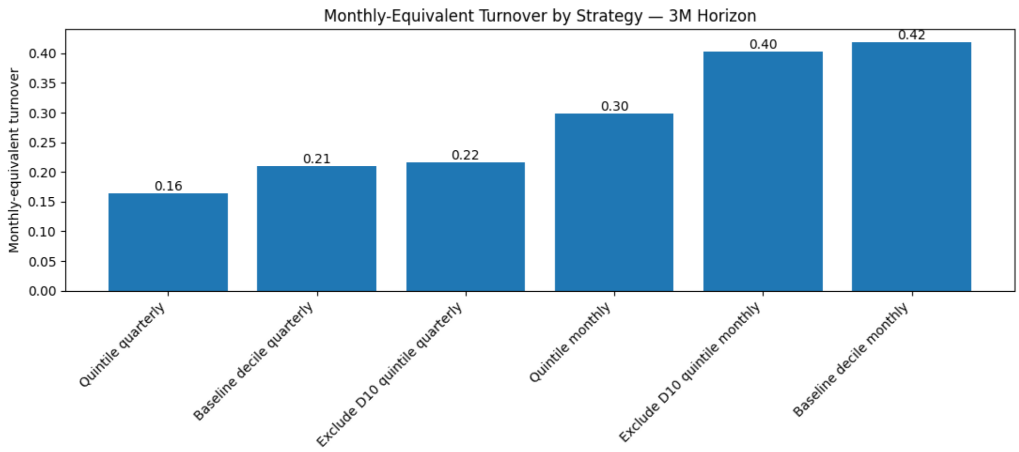
Notes: This table reports the sensitivity of 3-month net returns and Newey–West t-statistics to alternative annual borrow-fee assumptions.

Figure 28: Borrow-Fee Sensitivity of 3-Month Net Returns



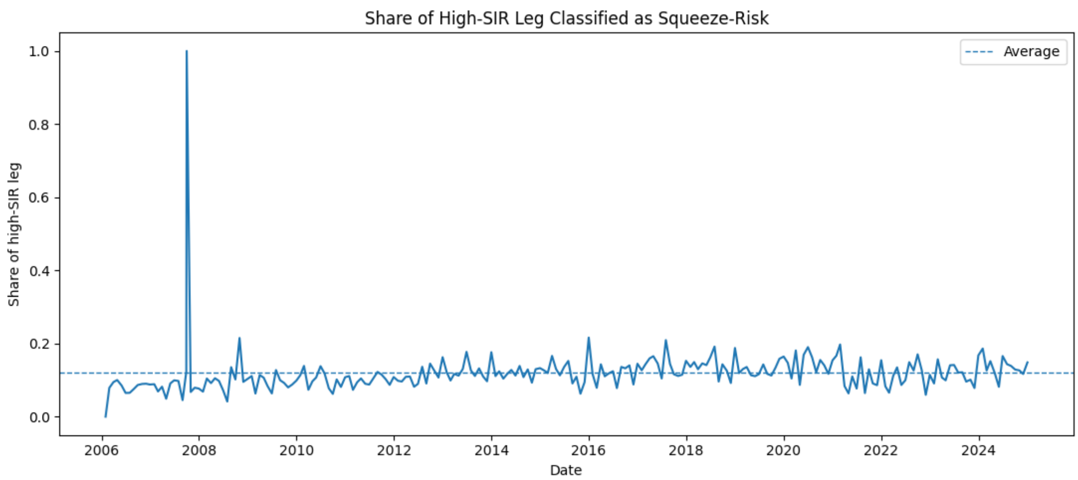
Notes: This figure shows how 3-month net returns change across alternative annual borrow-fee assumptions.

Figure 29: Monthly-Equivalent Turnover by Strategy at the 3-Month Horizon



Notes: This figure compares monthly-equivalent turnover across portfolio constructions. Quarterly turnover is divided by three to make it comparable with monthly rebalancing.

Figure 30: Share of the High-SIR Leg Classified as Squeeze-Risk



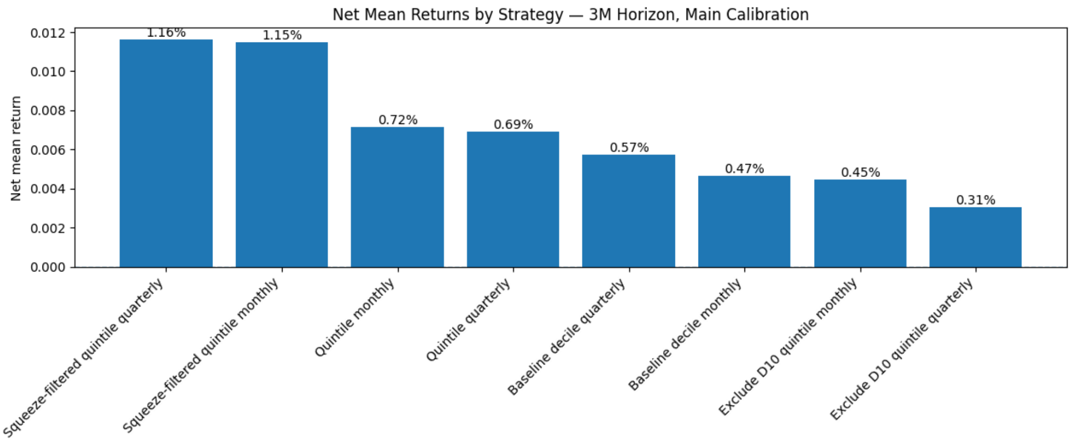
Notes: This figure plots the monthly share of stocks in the high-SIR short leg classified as squeeze-risk. A stock is classified as squeeze-risk if it is simultaneously in the top quintile of short interest, days-to-cover, and past one-month returns.

Figure 31: Post-Cost Annualized Net Returns Across Holding Horizons: Long-Short Strategies

Strategy	1m	3m	6m	12m
Long-short: Baseline decile monthly	0.0216 (0.738)	0.0188 (0.705)	0.0012 (0.045)	-0.0234 (-0.891)
Long-short: Exclude D10 quintile monthly	0.0157 (0.792)	0.0181 (1.034)	0.0194 (1.084)	0.0119 (0.686)
Long-short: Quintile monthly	0.0324 (1.246)	0.0290 (1.270)	0.0201 (0.864)	0.0008 (0.036)
Long-short: Squeeze-filtered quintile monthly	0.0480* (1.814)	0.0467** (2.023)	0.0392* (1.654)	0.0194 (0.835)
Long-short: Squeeze-filtered quintile quarterly	0.0879* (1.877)	0.0474* (1.700)	0.0288 (1.153)	0.0164 (0.676)

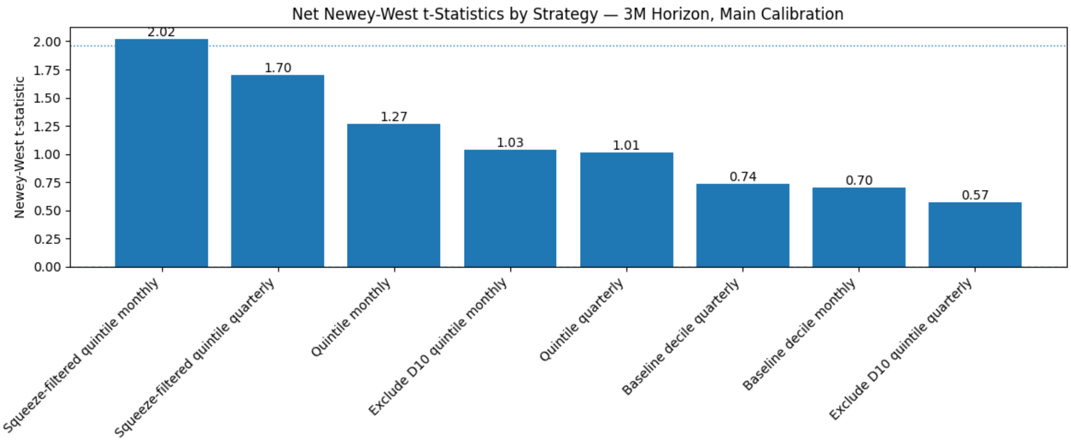
Notes: This figure reports annualized post-transaction-cost returns across holding horizons for long-short strategies under the main calibration. Newey–West t-statistics are reported in parentheses. Returns incorporate long-side trading costs, short-side trading costs, and borrow-cost assumptions. Significance levels are denoted by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Figure 32: Net Mean Returns by Strategy at the 3-Month Horizon under the Main Cost Calibration



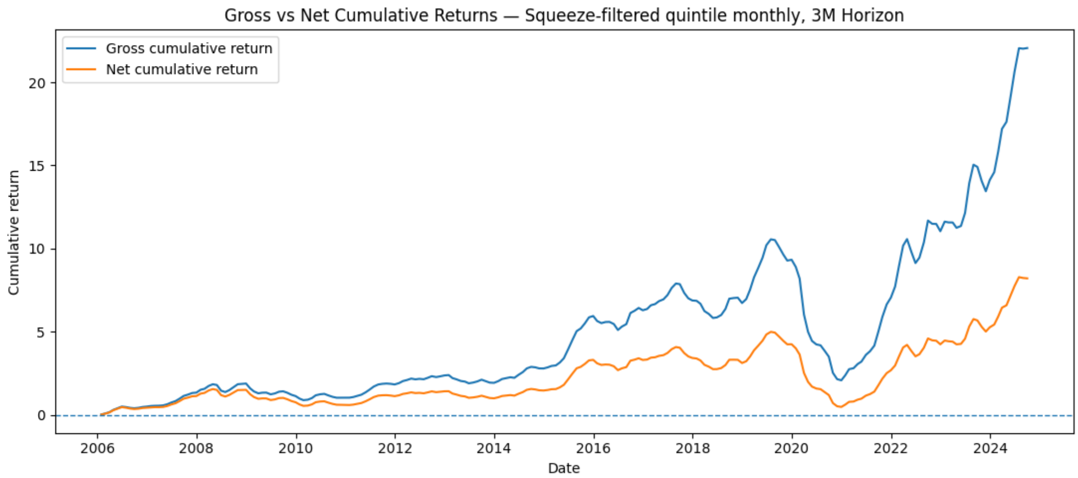
Notes: This figure reports 3-month net mean returns after transaction costs under the main calibration. The comparison includes the baseline decile, quintile, exclude-D10, and squeeze-filtered strategies under both monthly and quarterly rebalancing. Net returns incorporate long-side trading costs, short-side trading costs, and borrow-fee assumptions.

Figure 33: Net Newey–West t-Statistics by Strategy at the 3-Month Horizon under the Main Cost Calibration



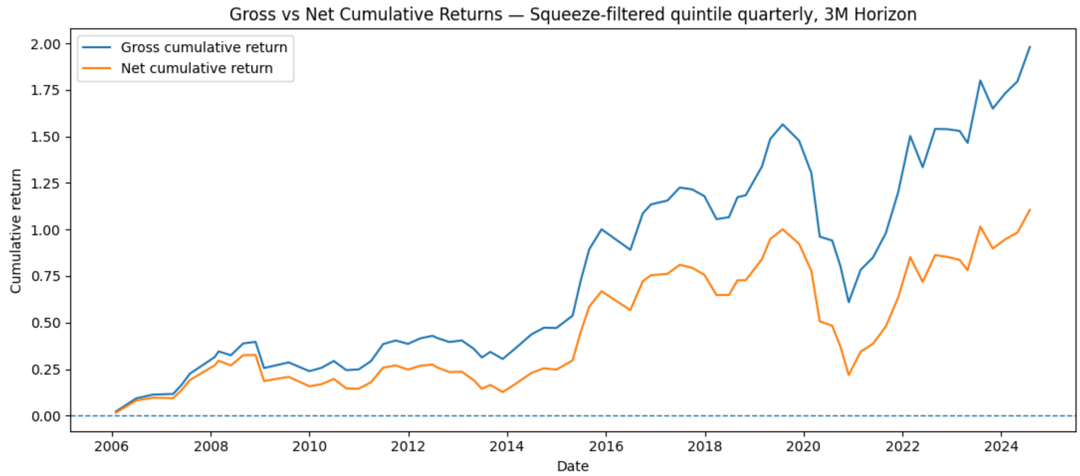
Notes: This figure reports Newey–West t-statistics for 3-month net returns under the main transaction-cost calibration. The dotted line marks the conventional $t = 1.96$ significance threshold.

Figure 34: Gross versus Net Cumulative Returns: Squeeze-Filtered Monthly Quintile Strategy



Notes: This figure compares gross and net cumulative returns for the squeeze-filtered monthly quintile strategy at the 3-month horizon under the main transaction cost calibration.

Figure 35: Gross versus Net Cumulative Returns: Squeeze-Filtered Quarterly Quintile Strategy



Notes: This figure compares gross and net cumulative returns for the squeeze-filtered quarterly quintile strategy at the 3-month horizon under the main transaction cost calibration.

A.5 Long-Only Enhanced Portfolio Results

Figure 36: Post-Cost Annualized Net Returns Across Holding Horizons

Strategy	1m	3m	6m	12m
Long-only: Benchmark EW	0.1885*** (3.745)	0.1705*** (4.916)	0.1499*** (4.551)	0.1362*** (3.994)
Long-only: Exclude high SIR & low MOM	0.1953*** (3.917)	0.1761*** (5.156)	0.1532*** (4.737)	0.1374*** (4.116)
Long-only: Exclude high SIR OR low MOM	0.1117*** (2.688)	0.1150*** (3.427)	0.1144*** (3.496)	0.1237*** (3.564)
Long-only: Exclude top 20% SIR	0.1193*** (2.706)	0.1175*** (3.334)	0.1188*** (3.437)	0.1287*** (3.706)

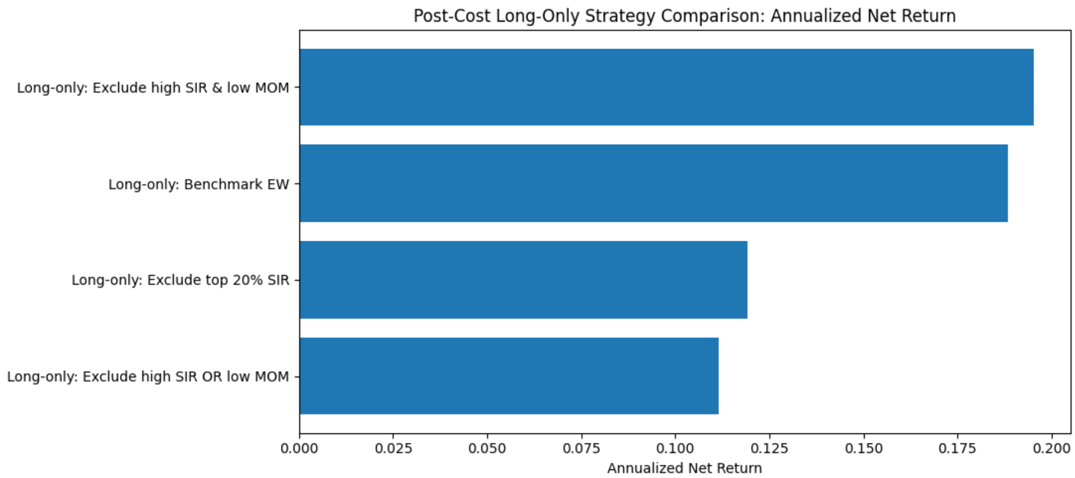
Notes: This figure reports annualized post-transaction-cost returns across holding horizons. Newey–West t-statistics are reported in parentheses. Long-only portfolios apply only long-side trading costs, assuming 25 basis points applied to realized turnover. Significance levels are denoted by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Figure 37: Post-Cost Sharpe Ratios Across Holding Horizons

Strategy	1m	3m	6m	12m
Long-only: Benchmark EW	0.7936	0.7857	0.5903	0.4822
Long-only: Exclude high SIR & low MOM	0.8323	0.8228	0.6107	0.4916
Long-only: Exclude high SIR OR low MOM	0.5974	0.6767	0.6378	0.5810
Long-only: Exclude top 20% SIR	0.6046	0.6539	0.6177	0.5675

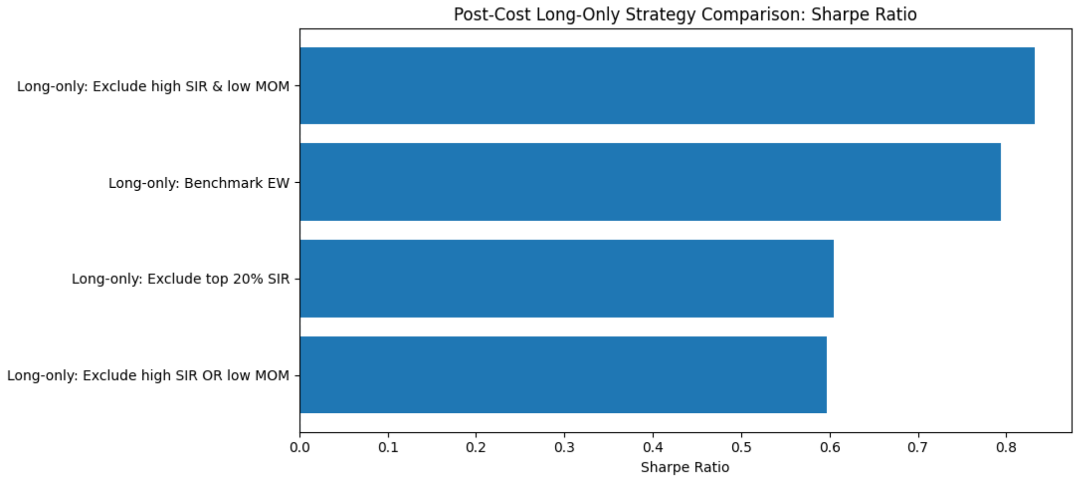
Notes: This figure reports post-transaction-cost Sharpe ratios across holding horizons for long-only portfolios. Sharpe ratios are computed using annualized net returns and annualized volatility after applying long-side trading costs of 25 basis points to realized turnover.

Figure 38: Post-Cost Long-Only Strategy Comparison: Annualized Net Returns



Notes: This figure compares annualized post-transaction-cost returns across long-only implementations. The benchmark portfolio is an equal-weighted portfolio of all eligible stocks. Alternative specifications exclude stocks with high short interest and/or low momentum. Transaction costs assume 25 basis points applied to realized turnover.

Figure 39: Post-Cost Long-Only Strategy Comparison: Sharpe Ratios



Notes: This figure compares post-cost Sharpe ratios across long-only implementations. Sharpe ratios are computed using annualized net returns and annualized volatility after transaction costs.