Segmented short sellers and predictable market returns[☆]

Job Market Paper

Arseny Gorbenko[*]

School of Banking & Finance, University of New South Wales, Sydney, NSW 2052, Australia

Current version: October 2020

Abstract

Short sellers, in aggregate, deviate from a well-diversified portfolio and concentrate their holdings in a limited segment of stocks with few short sale constraints. I show that this "segmentation" behavior effectively confines the set of stocks in which short sellers' systematic information can be incorporated into prices, leading to a market-wide underreaction to this information. As a result, short interest aggregated across stocks (negatively) predicts market returns. This predictability pattern holds in 30 out of 32 countries examined but is stronger in countries where short sellers hold more concentrated portfolios. Furthermore, short interest in the most heavily shorted stocks predicts returns on other stocks, signifying slow diffusion of systematic information to stocks beyond the short sellers' focus. Overall, the evidence indicates that short sellers' segmentation contributes to the persistence of predictable market returns, thus increasing market inefficiencies.

JEL classification: G12, G14, G15

Keywords: Short selling, return predictability, segmentation, informed investors

---

[☆] I thank Francisco Barillas, Pedro Barroso, Oleg Chuprinin, Carole Comerton-Forde, Jerry Parwada, Luis Goncalves-Pinto, Frank Weikai Li, Konark Saxena, Breno Schmidt, Rik Sen, Jianfeng Shen, Hang Wang, Xuewu Wang (discussant), and Qifei Zhu as well as conference and seminar participants at the 2020 International Risk Management Conference (IRMC), 2020 Econometric Research in Finance (ERFIN) Workshop, 2020 UNSW-UniMelb Market Microstructure Workshop, 2020 International Conference on Derivatives and Capital Markets (ICDCM), and Tianjin University and the University of New South Wales for their helpful comments. All errors are my own.

[*] *E-mail address*: a.gorbenko@unsw.edu.au

1. Introduction

Existing literature supports the idea that short sellers are rational, well-informed investors.[1] Their activity facilitates the incorporation of negative information into prices, thus promoting price efficiency (Boehmer and Wu, 2013). However, when short sale constraints bind, short interest cannot reach high enough levels to correct overpricing quickly, leading to a persistent negative association between short interest and future stock returns (e.g., Figlewski, 1981; Asquith, Pathak, and Ritter, 2005). Rapach, Ringgenberg, and Zhou (2016) show that this association also holds at the market level, that is, short interest aggregated across stocks negatively predicts market returns. However, the market-level pattern does not mechanically follow from the dynamics at the stock level. Short interest predicts returns only in hard-to-borrow stocks (e.g., Asquith et al., 2005), which account for 1% of the U.S. market value (D'Avolio, 2002). It is unlikely that these stocks could drive return predictability of the whole market. Thus, market-level predictability is a puzzle that requires further investigation.

This puzzle has two dimensions. First, it is unclear how the aggregation of short interest across stocks can produce a powerful market predictor. Second, the reason why short sellers do not eliminate predictable market returns is not obvious. They could have incorporated their negative information into market prices via index futures or other index securities, which are cheap and easy to short. But they are not currently doing it; otherwise, there would be no return predictability in the first place. In this paper, I focus on the first dimension of the puzzle and propose and test the mechanism that can partially explain why aggregate short interest predicts market returns.[2]

[1] For example, short sellers anticipate future stock returns (e.g., Boehmer, Jones, and Zhang, 2008; Diether, Lee, and Werner, 2009a), negative corporate events (e.g., Christophe, Ferri, and Angel, 2004; Christophe, Ferri, and Hsieh, 2010; Karpoff and Lou, 2010), and aggregate cash flow news (Rapach, Ringgenberg, and Zhou, 2016).

[2] I leave a thorough examination of the second dimension of the puzzle to future research. However, in Section 6, I discuss factors that could discourage short sellers from conducting informed trading in market index securities. These factors could be reluctance to bear market risks, capital constraints, or behavioral biases.

My explanation is based on two presumptions about short sellers: 1) that they possess systematic information and 2) that, in aggregate, they trade in a limited segment of stocks. The first presumption must be true; otherwise, short interest aggregation across stocks will consolidate only idiosyncratic information irrelevant for predicting market returns. The second presumption builds on the empirical observation that high short interest levels in stocks are persistent (Asquith et al., 2005), suggesting that short sellers focus on the same stocks over time. I call this propensity to trade in a limited set of stocks "short sellers' segmentation." The more tilted the aggregate short sellers' portfolio is toward particular stock segments, the more segmented the short sellers.

Short sellers' segmentation effectively limits the set of stocks in which their systematic information is incorporated into prices. Accordingly, other stocks reflect this information with a delay, generating a wide enough underreaction for aggregate short interest to predict market returns. Since short interest has a systematic component, this mechanism also implies that short interest in the most shorted stock segment is informative about returns on other stock segments.[3]

My empirical results support the above mechanism. I employ international data to link cross-country differences in the predictive power of aggregate short interest to short sellers' differential levels of segmentation across countries.[4] Consistent with my prediction, I show that aggregate short interest predicts market returns more strongly in countries where short sellers hold more concentrated portfolios. Moreover, I find that short interest in the most heavily shorted stocks predicts returns on other stocks, indicating a slow diffusion of short sellers' systematic information outside their target stock segments.

[3] Part of short interest's predictive power could arise from short sellers' reluctance to trade on their systematic information as trading on such information requires bearing substantial systematic risks. In this case, only some systematic information will be incorporated into stock prices. Without short sellers' segmentation, however, this partial reflection of information will occur in a larger number of stocks. Therefore, in any case, the segmentation mechanism that I propose strengthens the predictive ability of aggregate short interest.

[4] Investigating this link in a single country is problematic because segmentation has little time-series variation.

I start by investigating the drivers of short sellers' segmentation through the lens of their industry concentration. Industry concentration is a natural choice to track segmentation because active investors often concentrate their holdings in industries where they have informational advantages (Kacperczyk, Sialm, and Zheng, 2005). However, part of this concentration may be driven by exogenous cross-sectional differences in short sale constraints or valuation.[5] For example, short sellers should be more active in industries that feature more stocks with an abundant supply of lendable shares or more stocks that tend to underperform. I compute two measures of short sellers' industry concentration (relative to the market) and confirm that, in aggregate, they tend to short industries with fewer short-sale-constrained stocks and with more stocks prone to overpricing. Specifically, short sellers tilt their portfolios toward industries comprised of stocks with a large supply of lendable shares, low lender concentration, and high idiosyncratic volatility.

From an international perspective, some countries have greater cross-sectional differences in short sale constraints and valuation than others. Accordingly, short sellers' industry concentrations vary across countries. The segmentation mechanism implies that these variations can generate differences in the predictive abilities of aggregate short interest across countries.

I next verify that short interest predicts market returns internationally and that its predictive power exhibits large cross-country variations. Following Rapach et al. (2016), I construct the short interest index (*SII*) by averaging, detrending, and standardizing weekly short interests across stocks within each of the 32 countries in my sample. I find that *SII* is negatively associated with future market returns in 30 countries, and in 20 countries, this association is statistically significant.

[5] The role of short sale constraints in this context is distinct from the one discussed in the prior literature, which examines how binding short sale constraints affect the *existing* short-selling activity. In this paper, short sale constraints (which do not need to be binding) discourage *future* short-selling activity by reducing incentives to discover information right now. This disincentive emerges because it is costlier for short sellers to trade a more constrained stock once they discover negative information on this stock.

I emphasize that these results are not driven by the absence of financial instruments that permit market-level short selling outside the United States: in 29 countries from my sample, short sellers have cheap-to-trade market index futures to correct market prices. On average, a one-standard-deviation increase in *SII* leads to a 0.60% (2.03%) lower market return in the next month (quarter). The economic impact of *SII* on market returns varies substantially across countries. For example, a one-standard-deviation increase in *SII* predicts a 1.96% lower next-month market return in Greece and a 0.78% higher next-month market return in Poland.[6]

I further link the cross-country variation in the predictive power of *SII* to variations in short sellers' segmentation. I find that countries with above-median segmentation (i.e., where short sellers hold more industry-concentrated portfolios) display a 60% stronger association between *SII* and next-month market returns. This segmentation effect gradually decays with the return horizon, suggesting that short sellers' systematic information is slowly spreading outside the stock segments in which they trade more intensively.

To more precisely demonstrate the mechanism behind the segmentation channel, I focus on cross-stock return predictabilities since the channel implies slow diffusion of systematic information from the more- to the less-shorted stocks. I split stocks into three groups based on their weekly short interest and examine the relation of an average short interest in one group to average future returns in the other groups. Consistent with the segmentation mechanism, short interest in the most shorted stocks predicts returns on other stocks, after accounting for the other stocks' own short interest and lead-lag cross-correlations in returns. I also find that short interest in the most shorted stocks predicts downward revisions in analyst earnings forecasts and negative

[6] The results are robust to using several ways to control for the Stambaugh (1999) bias, to using non-detrended *SII* and alternative weighting schemes to compute *SII*, and to using days-to-cover ratio instead of short interest as a measure of aggregate short-selling activity. The predictive power of *SII* survives outside economic recessions, thus alleviating a concern that the predictive power of *SII* lacks temporal stability (Priestley, 2019).

earnings surprises in other stocks, although this predictability is not very strong. These results alleviate concerns that my findings on cross-country segmentation effects are driven by an omitted country-specific characteristic. It would be difficult for such a characteristic to explain a consistent global cross-stock return predictability pattern that I document.

The information already contained in short interest could be used profitably in market index futures, which are cheap to short and available in 29 out of 32 countries in my sample. I construct a trading signal based on historical short interest data and find that a strategy that sells market index futures following the signal delivers an average monthly (quarterly) return of 0.91% (3.21%).[7] This result implies that other drivers of persistent predictable returns, such as time-varying aggregate risk premium and binding short sale constraints, are unlikely to explain my findings since they imply that investors cannot exploit this predictability.

Nevertheless, I conduct additional tests and show that these drivers cannot subsume the predictive ability of aggregate short interest. I find that short sale regulations (uptick rules, naked short sale bans, and centralized equity lending) and funding constraints exert little influence on the predictive power of *SII*. Popular predictors of the time-varying aggregate risk premium, measures of market sentiment, and investor disagreement also do not offset the predictive ability of *SII*.

My paper makes several contributions to the literature on short selling. First and foremost, I find a new factor—short sellers' segmentation—that strengthens a negative relation between short interest and future market returns. Importantly, this factor is absent in a similar stock-level relation, but its effect can partially explain why aggregation of short interest across stocks results in a strong market predictor.[8] Furthermore, I show that historical short interest data predict returns on market

[7] This trading signal has no "look-ahead" bias as I use only data available at the time of constructing the signal.
[8] At the stock level, the persistent negative relation between short interest and future returns is traditionally attributed to short sale constraints and investor disagreement (e.g., Miller, 1977; Harrison and Kreps, 1978; Scheinkman and Xiong, 2003; Blocher, Reed, and van Wesep, 2013).

index futures, indicating that investors do not use these data to correct market prices. At the stock level, investors also do not always utilize available data to trade on predictable return patterns (e.g., Wang, Yan, and Zheng, 2020). However, the stock-level behavior is likely to be driven by binding short sale constraints (see, e.g., Jones and Lamont, 2002) that are virtually absent in index futures. What drives such market-level behavior is an interesting question that I leave to future research. Finally, I corroborate previous findings of Rapach et al. (2016) out of sample and demonstrate that aggregate short interest predicts market returns in many markets outside the United States.

More broadly, this paper contributes to the literature on aggregate market-level inefficiencies. Baker and Wurgler (2007), Baker, Wurgler, and Yuan (2012), and Greenwood and Shleifer (2014), among others, have found that investor irrationalities, such as sentiment and extrapolative expectations, can bias market prices. On the other hand, rational investors may not fully correct market-level mispricing since they have more incentives to focus on eliminating micro rather than macro inefficiencies (Glasserman and Mamaysky, 2019; Gârleanu and Pedersen, 2020). My findings suggest that rational investors' segmentation can exacerbate market-level inefficiencies by restraining the set of stocks where systematic information is partially incorporated into prices.

2. Theoretical and empirical background

A large literature has investigated the relation between short interest and future returns at the stock level. In this section, I discuss how the findings from this literature relate to the market-level phenomenon that I study.

A number of papers document a negative association between the amount of short sales and future returns at the stock level (e.g., Figlewski, 1981; Aitken, Frino, McCorry, and Swan, 1998; Boehmer, Jones, and Zhang, 2008; Diether, Lee, and Werner, 2009a; Boehmer, Huszár, and

Jordan, 2010). However, these results are cross-sectional and hence do not necessarily hold in time series. To illustrate this point, consider the finding of Boehmer et al. (2010) that stocks with high short interest tend to have low future returns, while stocks with low short interest tend to have high future returns. These results indicate that some stocks in the economy are overpriced, while others are underpriced. Therefore, the aggregate effect of short interest (over all stocks) on market returns is unclear. Engelberg, McLean, Pontiff, and Ringgenberg (2019) demonstrate that most stock-level predictors perform poorly as aggregate market predictors, suggesting that cross-sectional predictors largely contain idiosyncratic information about future returns.

On the theoretical side, short interest, which represents short-selling demand, relates to low future returns via short sale constraints and disagreement among investors. A combination of these two factors can inflate prices: optimistic investors would buy stocks from less optimistic investors, creating a demand pressure, while short sale constraints would prevent pessimistic investors from offsetting this effect (e.g., Miller, 1977; Harrison and Kreps, 1978; Scheinkman and Xiong, 2003; Blocher, Reed, and van Wesep, 2013). Inflated prices should eventually decrease as disagreement among investors narrows; this effect also attracts short sellers, leading to a negative association between short interest and future returns. Atmas, Basak, and Ruan (2020) show that short interest negatively predicts returns only in hard-to-short stocks since other stocks are less prone to be overpriced in the first place. This theoretical prediction finds empirical support in Figlewski and Webb (1993), Asquith et al. (2005), and Beneish, Lee, and Nichols (2015), who show that the predictive power of short interest largely disappears in stocks with non-binding short sale constraints (optionable stocks, stocks with high institutional ownership and low lending fees).

Hypothetically, a combination of short sale constraints and investor disagreement at the market level can also explain why short interest predicts market returns. However, the availability of cheap-to-short market index instruments, such as futures, options, and ETFs, undermines this explanation. Without short sale constraints, short sellers can quickly incorporate their negative views into prices, so that dispersion of beliefs among investors cannot generate overpricing and return predictability. There may be other limits to arbitrage that can prevent bearish investors from correcting inflated market prices. For example, funding constraints can push short sellers to close their positions, creating a deleveraging risk (Richardson, Saffi, and Sigurdsson, 2017). More broadly, short sellers may not want not to correct market-level inefficiencies because this activity requires bearing substantial systematic risks, which short sellers may want to avoid.

I note that stock-level short sale constraints can partly explain why aggregate short interest predicts market returns. These constraints prevent stock prices from fully adjusting to negative information, leading to an underreaction to this information at the aggregate level. However, most stocks are easy and cheap to borrow and therefore should be more efficiently priced. D'Avolio (2002) categorizes around 10% of stocks in the United States as special, while Asquith et al. (2005) find that only 5% of stocks experience a shortage of lending supply. These stocks are generally small and contribute little to the market portfolio. For example, D'Avolio (2002) estimates that hard-to-borrow stocks in aggregate account for around 1% of the U.S. market value. However, short sale constraints are more severe outside the United States (Boehmer, Huszár, Wang, and Zhang, 2018), and hence can make a higher contribution to the predictive ability of short interest.

In this paper, I investigate how the tendency of short sellers to focus on specific stock segments—short sellers' segmentation—can affect a negative association between short interest and market returns. I argue that the segmentation strengthens a negative association between short interest and future market returns by confining the set of stocks in which short sellers' systematic information can be incorporated into prices. There is evidence that short sellers exploit systematic information in trading. For example, Huszár, Tan, and Zhang, (2017) show that they use industry information, which contains a large systematic component (Hong, Torous, and Valkanov, 2007). However, to the best of my knowledge, there are no papers exploring short sellers' segmentation and its implications for asset prices. I elaborate on potential drivers of short sellers' segmentation in Section 3.3, and Section 5.1 empirically links segmentation to predictable market returns.

3. Data, main variables, and institutional details

My data sample covers 32 countries (23 developed and 9 emerging markets) for the period from July 1, 2006 to December 31, 2016.[9] I collect data from three sources. U.S. stock-level data on prices, returns, and the number of shares outstanding comes from CRSP. Datastream provides similar data for other countries as well as data on market indexes, index futures and options, ETFs, and government bonds. Markit provides international securities lending data for stocks and ETFs.

3.1. Sample selection

The filtration and merging of all three data sets follow Griffin, Kelly, and Nardari (2010), Hou, Karolyi, and Kho (2011), and Boehmer et al. (2018).[10] For most counties, I select only common stocks traded on the country's major exchange (the exchange with the highest number of traded stocks). Several exchanges are analyzed for Japan (Tokyo and Osaka), Korea (KSE and KOSDAQ), Singapore (Mainboard and Catalist), Spain (Madrid and Mercado Continuo Español), Taiwan (TWSE and TWO), and the United States (NYSE, Amex, and Nasdaq).

[9] I start with a sample of 38 countries as in Boehmer et al. (2018). However, I exclude Chile, China, Hungary, Indonesia, Malaysia, and the Philippines from my sample because they have too few stocks with non-missing short-selling data after I apply all filters.
[10] See Appendix A in Boehmer et al. (2018) for details.

I use the raw primary local market index returns, which account for dividend distributions, from Datastream to measure country-level market returns (Table A1 of Appendix A lists the market indexes used in the paper). Using raw returns allows for the inclusion of more countries in the analysis since risk-free interest rates are available for a sufficiently long period only for 22 out of 32 countries in my sample. Nevertheless, all results in the paper hold if I use excess market index returns instead of raw returns. I apply standard filters from Ince and Porter (2003) and Griffin et al. (2010) to eliminate errors and outliers when calculating returns.

I retrieve short-selling variables from the buy-side Markit database, which covers about 90% of securities lending transactions in developed countries (Gargano, Sotes-Paladino, and Verwijmeren, 2019); these countries compose the majority of my sample. Saffi and Sigurddson (2011) provide a detailed overview of the Markit database and international stock lending markets. Since short selling requires borrowing of the securities, stock lending transactions provide good proxies to measure the amount of stocks being sold short.

Markit reports daily values for the number of stocks currently on loan (*OnLoan*) and the number of stocks available for lending (*Lendable*), among other variables. To eliminate data errors and select stocks with sufficient variation in short sales throughout my sample period, I apply the following filters. First, I delete stocks that have missing or zero *OnLoan* values for the entire sample period. Second, I set *OnLoan* to equal zero if the amount of stocks on loan is missing while the amount of lendable stocks is not. Third, if Markit reports several observations for the same stock-day, I choose the observation with the highest amount of lendable stocks as it most closely corresponds with *Lendable* values on surrounding days.

3.2. Short selling across countries

To gauge the level of short-selling activity in each country, I aggregate stock-level information following Rapach et al. (2016). The idea is to aggregate short sales across stocks such that their idiosyncratic component vanishes while the common market component on which short sellers trade remains. First, I calculate the short interest in each stock, defined as the number of stocks on loan divided by the number of shares outstanding. Then, for each country, I calculate the daily equal-weighted average short interest (*EWSI*) across all stocks and detrend the resulting time series to remove the uninformative variation in *EWSI* that is not related to the short sellers' changing beliefs.[11] Specifically, I run the following time-series OLS regressions within each country:

$$Ln(EWSI)_t = \alpha + \beta \cdot t + \varepsilon_t, \tag{1}$$

where $Ln(EWSI)_t$ is the natural log of *EWSI* on day t. Following Rapach et al. (2016), I use the fitted residuals from this regression to measure aggregate short-selling activity and denote this measure the short interest index (*SII*). *SII* is further standardized to have a standard deviation of one (by construction, it also has a mean of zero) within each country. To reduce potential noise in the daily data, I calculate the average *SII* for each week and employ this weekly series in the subsequent regression analysis. I exclude a weekly *SII* from the sample if there are less than four trading days in a week.

Table 1 reports descriptive statistics for 32 countries in the final sample. The number of unique stocks included in the sample varies from 36 (Portugal) to 6,494 (the United States). The average daily short interest across stocks (*EWSI*) in most countries does not exceed 1%, whereas the U.S. *EWSI* is much higher at 3.55%. Short sellers trade less outside the United States, probably because

[11] Such variation can occur, for example, because of the increasing number of lendable shares due to the development of the stock lending market over time (Prado, Saffi, and Sturgess, 2016). In robustness tests (Section 4.2), I use alternative weighting schemes in aggregating short interest across stocks.

they face more short sale constraints. For example, an average (median) annualized stock lending fee outside the United States is 3.55% (2.36%) compared to 1.30% (0.38%) in the United States (untabulated).[12] More severe short sale constraints could make short sales more informative by discouraging liquidity-motivated trades (Diamond and Verrecchia, 1987), leading to a more widespread negative association between short interest and future stock returns. Boehmer et al. (2018) lend empirical support for this idea; this effect could lead to a stronger ability of aggregate short interest to predict market returns outside the U.S. market as discussed in Section 2.

In most countries, short sellers can sell the whole market by trading market index futures, options, or ETFs. The choice of an instrument largely depends on its availability and liquidity. The most natural first choice in most countries is index futures because they are available and usually very liquid. Sutcliffe (2006) compares trading in index futures with that in stocks and concludes that index futures are easier to short, are more liquid, and are associated with lower bid-ask spreads and trading commissions. Index options and ETFs are less widespread and are available only in 18 and 15 countries from my sample, respectively. Yet, these securities are also liquid and cheap to short (Gastineau, 2010; Han and Li, 2017). For example, the average lending fee of the S&P 500 ETF (SPY) during my sample period is only 0.44% (vs. 1.30% for an average U.S. stock).

[Table 1 here]

Short-selling activities across countries could correlate if they are driven by a global factor. To check this possibility, I plot cross-country correlations between daily *SII*s, which are standardized within each country. In Fig. 1, each box and whisker plot shows the range of correlations between a given country's *SII* and *SII*s in other countries. A cross (a horizontal line) in each box indicates the mean (median) correlation between the given country's *SII* and *SII*s in other countries. Most

[12] The similar average global and U.S. lending fees in excess of ocal risk-free rates are 1.17% and 0.43%, respectively.

cross-country correlations are positive, but some countries' *SII*s exhibit mostly negative correlations with other countries' *SII*s (Brazil, Greece, Poland, Russia, and Taiwan). The mean cross-country correlation between *SII*s equals to a modest 0.20. Overall, these correlation patterns indicate that short-selling activities in each country capture mostly local information.

[Fig. 1 here]

3.3. Short sellers' segmentation across countries

One of the paper's major objectives is to estimate how the level of short sellers' segmentation affects the predictive power of aggregate short interest. In this section, I construct two segmentation measures and examine potential factors that drive short sellers' segmentation.

It is important to stress that the decision to concentrate on a limited segment of stocks, instead of holding a diversified portfolio as prescribed by traditional asset pricing, can be optimal if investors have informational advantages in those segments. Van Nieuwerburgh and Veldkamp (2009, 2010) show theoretically that increasing returns to scale in learning prompt investors to hold concentrated portfolios. Empirically, Kacperczyk et al. (2005) documents that many U.S. mutual funds hold industry-concentrated portfolios (relative to the market), while an international study of Choi, Fedenia Skiba, and Sokolyk (2017) find that institutional investors with higher learning capacities (i.e., more skilled investors) hold more industry-concentrated portfolios.

I follow the literature and examine short sellers' segmentation through the lens of their industry concentration.[13] However, unlike previous papers, I focus on industry concentration of the *aggregate* short sellers' portfolio. It is an important distinction because I am interested in whether short sellers, in aggregate, tend to segment in specific industries since this collective segmentation

[13] Throughout the paper, I use the terms "short seller's segmentation" and "short sellers' industry concentration" interchangibly.

is what drives slow diffusion of systematic information from one stock segment to the other. Aggregate industry concentration can be driven by exogenous factors, such as cross-stock differences in limits to arbitrage or valuation. Specifically, some industries could feature stocks with fewer limits to arbitrage or more inclined to overvaluation; this effect would prompt short sellers to tilt their portfolios to these industries.

From an international perspective, some countries are likely to have bigger cross-stock differences in limits to arbitrage or (over)valuation than others. This cross-country variation would drive variation in short sellers' industry concentration, which I am going to use to study the impact of segmentation on the predictive power of aggregate short interest.

I construct two measures of short sellers' industry concentration. These measures draw from the literature on portfolio concentration (Kacperczyk et al., 2005) and active portfolio management (Cremers and Petajisto, 2009; Doshi, Elkamhi, and Simutin, 2015). The first measure, *IdustryConc*, is the Herfindahl-type concentration index defined as the sum of the squared deviations of the value weights of each industry *i* in the portfolio consisting of all short positions in a country at the end of week *t*, $w_{i,t}^{s}$, relative to the industry weights of the market portfolio, $w_{i,t}^{m}$:

$$IndustryConc_t = \sum_{i=1}^{n} \left(w_{i,t}^{s} - w_{i,t}^{m}\right)^2. \tag{2}$$

I adopt the most detailed level of industry classification, which corresponds to the 4-digit Standard Industry Classification Code (SICCD) in the United States and the Datastream Level 6 industrial classification number (INDG) in other countries. *IndustryConc* equals zero if the collective portfolio of all short sellers has the same industry weights as the market portfolio, and increases as this portfolio becomes more concentrated on a few industries.

Because of the squaring in the computation method, the above measure will generally assign a lower value to a country with a higher number of industries. This feature can introduce noise into the cross-country comparison of short sellers' segmentations. Thus, I construct an alternative measure, *IndustrySpec*, which aims to mitigate the above issue:

$$IndustrySpec_t = \frac{1}{2}\sum_{i=1}^{n}\left|w_{i,t}^{s} - w_{i,t}^{m}\right|. \tag{3}$$

IndustrySpec increases as short sellers start to bet more heavily on particular industries.

Both *IndustryConc* and *IndustrySpec* capture similar information, as indicated by their high correlation of 0.87. Most variation in both measures occurs across countries rather than over time. The average (median) cross-country standard deviations of *IndustryConc* and *IndustrySpec* over time are 0.16 (0.16) and 0.15 (0.14), respectively. The similar average (median) time-series standard deviations across countries are 0.09 (0.04) and 0.09 (0.08), respectively. It thus makes sense to focus more on cross-country variation in segmentation for further analysis.

Fig. 2 depicts the average values of *IndustryConc* and *IndustrySpec* in 32 countries during my sample period. Both segmentation measures exhibit significant cross-country variations and classify similar countries as the ones with the most and least segmented short sellers. For example, both measures classify Russia and Greece as the countries with the most segmented short sellers. However, there are some discrepancies between the measures as well. For instance, *IndustryConc* classifies the United States as the country with the least segmented short sellers, while *IndustrySpec* puts the U.S. short sellers on the 20th place by segmentation. The difference occurs because *IndustryConc* assigns lower ranks to countries with a large number of industries.

[Fig. 2 here]

Next, I examine how limits to arbitrage and propensity to overvaluation in stocks impact on aggregate industry concentration of short sellers. The most direct limit to arbitrage that affects short sellers is short sale constraints. I consider three types of such constraints that have been shown to have a material impact on short sellers: supply of lendable shares (Saffi and Sigurdsson, 2011), lender concentration (Kolasinski, Reed, and Ringgenberg, 2013), and lending fee variability also known as short-selling risk (Engelberg, Reed, and Ringgenberg, 2018). I also consider liquidity as a potential limit to arbitrage. On the valuation side, I look at idiosyncratic volatility, which has been shown to be associated with negative future returns (Ang, Hodrick, Xing, and Zhang, 2006; 2009), and book-to-market ratio.

I run the following panel regressions within each country:

$$\begin{aligned} Overweight_{i,t} &= \alpha + \beta_1 Lendable_{i,t-1} + \beta_2 LenderConc_{i,t-1} + \beta_3 ShortRisk_{i,t-1} \\ &+ \beta_4 FHT_{i,t-1} + \beta_5 IVol_{i,t-1} + \beta_6 BM_{i,t-1} + \beta_7 LogMCap_{i,t-1} + FE_t + \varepsilon_{i,t}, \end{aligned} \tag{4}$$

where $Overweight_{i,t}$ is the difference between value weights of industry *i* in the aggregate short sellers' portfolio and the market portfolio at the end of week *t*, $w_{i,t}^s - w_{i,t}^m$. All independent variables, except for *LogMCap*, are value-weighted averages for stocks in industry *i* at the end of week *t* (weekly, monthly, and yearly variables are estimated up to week *t–1*). *Lendable* denotes the number of stocks available for lending (standardized by the number of shares outstanding), *LenderConc* is the average weekly metric of lenders concentration from Markit, *ShortRisk* is the yearly lending fee variability, *FHT* is a monthly illiquidity measure from Fong, Holden, and Trzcinka (2017),[14] *IVol* is the monthly idiosyncratic volatility estimated from the model of Hou et al. (2011),[15] *BM* is the last year-end book-to-market ratio. *LogMCap* is the cumulative market capitalization of all stocks in industry *i* at the end of week *t–1*. I include *LogMCap* to check whether short sellers underweight industries that have larger contributions to the market portfolio. All regressions include week fixed effects. I standardize all independent variables within each country.

[14] Fong et al. (2017) show that their liquidity proxy, *FHT*, is one of the best percent-cost liquidity proxies in international markets when compared to high-frequency liquidity measures, such as effective spread. Nevertheless, my results remain qualitatively similar if I use an illiquidity measure from Amihud (2002).

[15] Hou et al. (2011) model consists of three global and three local factors: market factor, momentum, and cash-flow-to-price. The auhors show that their model outperforms other models in explaining the cross section of stock return in international markets.

Panel A of Table 2 reports country-specific results. In most countries, coefficients on *Lendable* are significantly positive, and coefficients on *LenderConc* are significantly negative, indicating that short sellers overweight industries with stocks that have a larger supply of lendable shares and less lender concentration. Other variables have less consistent signs of the coefficients across countries. Nevertheless, median coefficients across countries are consistent with short sellers avoiding short sale constraints and targeting overvalued stocks. A negative coefficient on *ShortRisk* suggests that bearish investors avoid stocks with high lending fee variability. A positive coefficient on *IVol* and a negative coefficient on *BM* suggest that short sellers tilt their portfolios toward industries that feature more potentially overvalued stocks. A negative coefficient on *LogMCap* indicates that short sellers underweight industries that have larger contributions to the market portfolio. In Panels B and C, I run pooled global panel regressions with country fixed effects. The results remain qualitatively similar.[16]

[Table 2 here]

Overall, these results indicate that, in aggregate, short sellers tend to hold (short) industries featuring less short-sale-constrained stocks and more stocks inclined to overvaluation.

4. Does short interest predict market returns globally?

Currently, the only evidence on the ability of short interest to predict market returns comes from the U.S. market (Rapach et al., 2016). The goal of this section is to investigate the global prevalence and cross-country variation of this predictability pattern.

[16] In Table A2 of Appendix A, I run similar regressions as in Eq. (4), but include standard deviation of analysts' EPS forecasts (from I/B/E/S) as an additional independent variable that proxies for investor disagreement (which can increases the stock's overvaluation as discussed in Section 2). The results remain qualitatively similar, with *Lendable* and *LenderConc* being the most important drivers of short sellers' industry concentration and investor disagreement having an expected positive influence on future short interest. I do not include my measure of investor disagreement in the main regression specification because the required data on analyst forecasts outside the United States are scarce.

4.1. Baseline results

I begin the analysis by running the following time-series regressions within each country:

$$RET_{t:t+n} = \alpha + \beta_1 SII_{t-1} + \beta_2 RET_{t-n:t-1} + \beta_3 US\ RET_{t-n:t-1} + \varepsilon_t, \tag{5}$$

where $RET_{t:t+n}$ is the cumulative market index return in a given country for *n* weeks starting from week *t*, SII_{t-1} is the average daily *SII* in country *i* for week *t–1*, $RET_{t-n:t-1}$ is the cumulative market index return for the past *n* weeks ending in week *t–1*, and $US\ RET_{t-n:t-1}$ is the cumulative S&P 500 index return for the last *n* weeks ending in week *t–1*. I use returns over 4-, 8-, and 12-week horizons, which approximately correspond to 1, 2, and 3 calendar months, respectively. Since U.S. market returns lead other countries' returns (Rapach, Strauss, and Zhou, 2013), I control for the lagged local and U.S. market returns in all regressions. I emphasize that *SII* is standardized to have a mean of zero and a standard deviation of one within each country; this facilitates a comparison of the results between countries. It is well known that Stambaugh (1999) bias and overlapping observations (e.g., Hodrick, 1992) complicate statistical inferences in Eq. (4). To address these complications, I use heteroskedasticity- and autocorrelation-robust Newey-West (1987) *t*-statistics with 12 lags. I multiply all coefficients by 10,000 to present them in basis points.

Table 3 reports the OLS estimates of β_1 in Eq. (5) for different return horizons. Columns (1)–(3) of Panel A in the upper section of the table shows that *SII* has the expected negative sign in 30 out of 32 countries at most horizons. In 20 countries, these coefficients are also statistically significant. These results are quite impressive given that my time series in each country spans only for around 11 years. The economic significance varies substantially across countries. For instance, a one-standard-deviation increase in *SII* predicts a 195.70 bps *lower* one-month market return in Greece and a 78.34 bps *higher* one-month market return in Poland.

To measure the average impact of aggregate short interest on market returns, I estimate the pooled version of Eq. (4) that imposes the restriction of homogeneous slope coefficients across countries. Even if this restriction does not hold exactly, pooled estimates can meaningfully measure average relation in the data (e.g., Hjalmarsson, 2010; Rapach et al., 2013).[17]

Columns (1)–(3) of Panel B report the results. On average, a one-standard-deviation increase in *SII* corresponds to a 60.37 bps lower market return in the following month (7.85% annualized). At the quarterly horizon, a one-standard-deviation increase in *SII* is associated with 203.34 bps lower market return (8.81% annualized). Predictive regressions produce R^2 statistics of 1.07% (4.81%) at the monthly (quarterly) horizon. These are large numbers, given that other well-known predictors of market returns (e.g., the dividend-price ratio, dividend yield, and net equity expansions) produce R^2 statistics rarely exceeding 0.5% (1%) at the monthly (quarterly) horizon (Rapach et al., 2016).

All estimates are statistically significant when I use Newey-West standard errors. However, these standard errors do not account for potential cross-country correlations due to exposure to common global market shocks. To address this issue, I report *t*-statistics based on Driscoll-Kraay (1998) standard errors that account for the general forms of cross-sectional dependence, heteroskedasticity, and autocorrelation in the error structure. All statistical inferences remain unchanged. Since Driscoll-Kraay standard errors produce more conservative statistical inferences, I continue using them in panel regressions in the rest of the paper.

Rapach et al. (2016) detrend the aggregate short interest to remove the strong upward long-term trend in the data. The concern is that daily detrending may remove more information than noise. I

[17] Since the pooled regression specification includes all countries, I drop $US\ RET_{t-n}$ from estimations. However, my results hold if I exclude the United States from the sample and keep using $US\ RET_{t-n}$. Although I do not use country fixed effects in pooled global panel regressions to avoid potential statistical biases arising from using persistent regressors (Hjalmarsson, 2010), my results remain unchanged if I do include them.

thus re-estimate Eq. (5) for the non-detrended measure of aggregate short-selling activity, *EWSI*. Consistent with the previous findings, columns (4)–(6) of Table 2 show that increasing *EWSI* is associated with decreasing future market returns.

[Table 3 here]

In Table 3, the economic and statistical significance of *SII* and *EWSI* increases with the return horizon (in annualized terms). This result is consistent with Wang et al. (2020) who show that shorting flows are more informative about longer-term returns. They argue that this pattern occurs because short-sellers trade mostly on long-term information, and hence the short-term returns reflect less of this information compared to the long-term returns. However, this pattern can also arise because of statistical biases that increase with the return horizon (e.g., Hodrick, 1992; Ang and Bekaert, 2007). I address this concern in the next section.

4.2. Robustness

In what follows, I discuss several robustness checks, which confirm the baseline findings from the previous section. I report these results in Tables A4–A6 of Appendix A.[18]

Like many popular predictors of market returns, *SII* is highly persistent with a weekly autocorrelation of 0.96. Persistence coupled with overlapping observations raise econometric concerns (e.g., Torous, Valkanov, and Yan, 2004; Ang and Bekaert, 2007). To address them, I use the Wald test developed by Kostakis, Magdalinos, and Stamatogiannis (2015). This test is robust to various forms of persistence (unit root, local-to-unit root, near stationarity, and stationarity) and is applicable to multivariate long-horizon predictive regressions. The key idea of the method is to construct an instrumental variable, whose degree of persistence is explicitly controlled.

[18] I do not tabulate all the results that I discuss in this section, but they are available upon request.

Table A4 of Appendix A reports the results of the above test for Eq. (5). The results are very similar to the ones in Table 3. Most notably, the magnitudes of the coefficients on the instrumented *SII* correspond well to the actual *SII* coefficients in Table 32. This fact reassures that the economic significance of my baseline results is unlikely to be driven by the persistence of *SII*.

Stambaugh (1999) describes how the correlation between the innovations to market returns and a lagged dividend yield can lead to biased regression results. Unlike dividend yield, short interest is not mechanically related to market returns. Nevertheless, to make sure that the Stambaugh (1999) bias does not drive my results, I conduct a simulation similar to Kothari and Shanken (1997), Ang and Bekaert (2007), and Yu (2011).[19] I run univariate regressions of market returns on the lagged *SII* to obtain the actual *SII* coefficients for each country. I assume that *SII* follows the AR(1) process and estimate a one-week autocorrelation coefficient for each country. Then, I draw the error terms (with replacement) from the joint empirical distribution of the residuals in the univariate regressions and the AR(1) model. These error terms are used to simulate an *SII* coefficient under the assumption that the actual *SII* coefficients estimated with the original data are "true" coefficients. I repeat this procedure 10,000 times. The difference between the actual *SII* and the average simulated *SII* represents an estimate of the Stambaugh (1999) bias.

Table A5 of Appendix A shows that the Stambaugh (1999) bias in time-series and panel regressions is small and rarely exceeds one basis point at all return horizons. On average, the Stambaugh (1999) bias reported in columns (4)–(6) represents less than 1% of the actual magnitudes of the *SII* coefficients in columns (1)–(3). To estimate the statistical significance of the coefficients, I run similar simulations as before under the assumption of null predictability (i.e., I set the "true" *SII* coefficients to zero). These simulations help to understand how likely one can

[19] Appendix B describes the simulation procedure in detail.

get similar actual *SII* coefficients and Newey-West *t*-statistics by chance. The results suggest that this is very unlikely. Both the actual magnitudes of the coefficients and their corresponding *t*-statistics are in line with the baseline findings in Section 4.1 and Table 3.

Priestley (2019) notes that the predictive power of *SII* in the United States disappears once he excludes the calendar year of 2008, which is characterized by a severe economic recession. To check whether recessions can subsume my results, I interact *SII* with a *Recession* dummy that equals one if a country experiences a technical recession. Table A6 of Appendix A shows *SII* continues to remain negative and significant in all regression specifications, indicating that the predictive ability of aggregate short interest survives outside economic recessions. However, a significant negative coefficient on the interaction, *SII*×*Recession*, suggests that *SII*'s predictive power increases in recessions, consistent with previous findings in the literature that market return predictors perform better around economic downturns (e.g., Henkel, Martin, and Nardari, 2011).

Hong, Li, Ni, Scheinkman, and Yan (2016) argue that the days to cover ratio (*DTCR*), calculated as the number of stocks on loan divided by the daily trading volume, is a superior measure compared to short interest. *DTCR* takes into account the average amount of time it takes to cover a short position, thus incorporating additional liquidity information. I construct the detrended *DTCR* index and use it instead of *SII* in regressions. The results remain qualitatively similar.

Finally, one could argue that using value-weighted short interest is more appropriate for predicting value-weighted market returns. However, this weighting scheme assigns a higher weight to large stocks that are characterized by less informative short sales (e.g., Desai, Ramesh, Thiagarajan, and Balachandran, 2002; Asquith et al., 2005). Thus, the equal-weighted short interest likely provides a more informative aggregate signal. Nonetheless, I construct *SII* using an alternative value-weighting scheme. I use the log of market capitalization to weight short interests in stocks for the construction of the short interest index. This weighting scheme mitigates placing too much emphasis on large stocks but recognizes their higher contribution to the value-weighted market returns. I run a similar test as in Eq. (5) with this alternative short interest index and find very similar results.

5. Why does short interest predict market returns?

The previous section demonstrates that short interest predicts market returns in many countries globally. This section investigates the potential reasons behind this pattern.

5.1. Short sellers' segmentation

My main conjecture is that short sellers' segmentation significantly increases the predictive power of *SII*. Specifically, short sellers' segmentation slows down the diffusion if their systematic information across stocks segments, generating a wide enough underreaction for the aggregate short interest to predict market returns. To test this conjecture, I examine the relation between short sellers' industry concentration and the predictive ability of short interest across countries.

Segmentation varies mostly across countries rather than over time (see Section 3.3). It thus can contribute to the cross-country differences in the predictive power of *SII* found in Section 4.1 and Table 3. If this conjecture is correct, we should observe a negative association between the level of short sellers' segmentation in a country and this country's coefficients on aggregate short interest. Fig. 3 provides graphical support for this conjecture. Specifically, the scatter plots between four-week *SII* coefficients (Panel A of Table 3) and short sellers' industry concentration measures (Fig. 2) across countries exhibit a negative trend. Consistent with the segmentation mechanism, this trend indicates that short interest has a stronger negative association with future market returns in countries where short sellers hold more concentrated portfolios.

[Fig. 3 here]

I also examine the empirical pattern from Fig. 3 in a regression framework. Each week, I split countries into two groups based on aggregate short sellers' segmentation: one where short sellers hold more industry-concentrated portfolios and another where they hold less concentrated portfolios. I want to find whether countries with more segmented short sellers have significantly different *SII*s. Thus I run the following global panel regressions:

$$\begin{aligned} RET_{i,t:t+n} = \alpha + \beta_1 SII_{i,t-1} + \beta_2 Segmented_{i,t-1} + \beta_3 SII_{i,t-1} \times Segmented_{i,t-1} \\ + \beta_4 RET_{i,t-n:t-1} + \varepsilon_{i,t}, \end{aligned} \tag{6}$$

where $Segmented_{i,t-1}$ is a dummy variable that equals one if short sellers in country *i* are characterized by the above-median segmentation (measured by *IndustryConc* or *IndustrySpec* described in Section 3.3) in week *t-1* across my sample countries, and zero otherwise. The interaction coefficient, β_3, has an intuitive interpretation: it shows how different is the short interest's predictive power in countries with more segmented short sellers.

Table 4 reports the results. Significantly negative coefficients on the interaction term in columns (1)–(3) indicate that high segmentation increases the predictive power of *SII*, consistent with my conjecture. The results are robust to using an alternative measure of short-selling activity (*EWSI*) in columns (4)–(6) and different segmentation measures in Panels A and B. A one-standard-deviation increase in *SII* has around 60% stronger effect on the next-month market return in countries with more segmented short sellers. The difference between countries gradually decreases to around 30% at a quarterly horizon. These dynamics suggest that stocks that receive less attention from short sellers gradually adjust to the information revealed in more heavily shorted stocks.

[Table 4 here]

The segmentation channel implies certain cross-segment relations between short interest and future return due to the slow diffusion of systematic information from the more to the less shorted stock segments. Specifically, short sellers realize their information advantage through trades exclusively in more shorted stock segments, so that short interest in these segments should predict returns on other stock segments since this information contains a systematic component. I thus split stocks into three groups based on their weekly short interest and examine the relation of an average short interest in one group and future returns in other stock groups.[20]

I begin by splitting stocks into terciles based on their short interests. I report statistics on these stocks in Table A7 of Appendix A. In an average country, more than 85% of stocks retain their assigned terciles for 4 consecutive weeks, and more than 70% of stocks retain their assigned terciles for 12 consecutive weeks. This pattern indicates that short sellers tend to trade more heavily in the same group of stocks, which most likely represent the stocks in which short sellers specialize. The most heavily shorted stocks have the highest market capitalization: their cumulative contribution to the market portfolio is, on average, around 60%. However, this contribution substantially varies across countries: from 17% in the United States to 86% in France. The contribution of the most shorted stocks to the aggregate short sellers' portfolio is more uniform across countries: from 64% in the United States to 99% in France and Singapore (untabulated). The least shorted stocks have a very small average short interest (*EWSI*) of 0.01% (vs. 2.00% in the most shorted stocks). In many countries, these stocks often exhibit no shorting during the week. Since *SII* drops observations with zero *EWSI* by construction, statistical inferences may be biased due to a small number of non-missing observations; I thus base my subsequent analysis on *EWSI*.

[20] Alternatively, I could split stocks by more and less shorted industries. However, in some countries, this spit would produce stock segments consisting of a small number of stocks. This could introduce additional noise to my analysis.

Consider the following regression model:

$$TercX\ RET_{i,t:t+n} = \alpha + \beta_1 Terc1\ EWSI_{i,t-1} + \beta_2 Terc2\ EWSI_{i,t-1} + \beta_3 Terc3\ EWSI_{i,t-1} + \beta_4 Terc1\ RET_{i,t-n:t-1} + \beta_4 Terc2\ RET_{i,t-n:t-1} + \beta_4 Terc3\ RET_{i,t-n:t-1} + \varepsilon_{i,t}, \tag{7}$$

where *TercX RET* (*TercX EWSI*) denotes the equal-weighted average return (short interest) on the portfolio consisting of stocks categorized into tercile *X* based on their weekly short interest.[21] The segmentation hypothesis implies that *Terc3 EWSI* has a significant and negative impact on future returns of stocks in other terciles. Simultaneously, short interest in other terciles should have minimal predictive power with respect to *Terc3 RET* because short sellers conduct very little trading in these stock groups and hence are unlikely to reveal any systematic information there. I note that I control for lagged returns in all regressions since non-synchronous trading across stocks can results in cross-stock return autocorrelations (see, e.g., Lo and MacKinlay, 1990). I standardize *TercX EWSI* within each county.

Table 5 reports the OLS estimates of β_1–β_3 in Eq. (7). Consistent with the segmentation channel, columns (1)–(6) show that short interest in the most shorted stocks (*Terc3 EWSI*) has a significant negative impact on future returns of the less shorted stocks (*Terc1 RET* and *Terc2 RET*) at all return horizons. Conversely, short interest in less shorted stocks have little effect on future returns in the most shorted stocks, as indicated by insignificant coefficients on *Terc1 EWSI* and *Terc2 EWSI* in columns (7)–(9).

[Table 5 here]

Notably, short interest in the most shorted stock group predicts its own returns, as indicated in columns (7)–(9), suggesting that short sellers restrain from trading on their systematic information even in stocks in which they specialize. Perhaps, this behavior is driven by the fact that trading on

[21] To conserve space, I focus on the pooled global panel regressions. Time-series regressions yield similar results.

this information requires bearing systematic risks. Gârleanu and Pedersen (2020) show theoretically that this fact disincentivizes active investors to correct market-level mispricing, so that they focus on cross-sectional mispricing instead. While correcting cross-sectional mispricing, investors can hedge their systematic exposure. For example, Huang, O'Hara, and Zhong (2020) demonstrate that investors buy stocks and short sell industry ETFs prior to positive earnings announcements to hedge against industry risks.

If short sellers dislike systematic risks, why do they acquire systematic information then? The likely reason is that, in practice, it is hard to obtain purely idiosyncratic information. For example, private information about a company's poor performance may arise due to more exogenous, systematic industry factors. As a result, short sellers would acquire information that has some systematic component, so that in aggregate, their trades would be able to predict market returns.

An alternative way to test the slow systematic information diffusion from the more to the less shorted stocks is to examine whether short interest in the most shorted stocks predicts earnings news in other stocks. Specifically, high short interest in the most shorted stocks indicates bad news for other market segments and therefore should be associated with more downward revisions in analyst earnings forecasts and more negative earnings surprises in those segments. To investigate this possibility, in Table A8 of Appendix A, I run regressions as in Eq. (7) but replace the dependent variable, a portfolio return, with the proportion of the stocks in the portfolio that experience downward revisions in analyst earnings forecasts or negative earnings surprises. The table shows that short interest in the most shorted stocks predicts more downward revisions and negative earnings surprises in less shorted portfolio terciles (*Terc1* and *Terc2*), although this effect is economically small and is not always statistically significant. Nevertheless, the overall pattern is consistent with the slow diffusion of systematic information outside the most shorted stocks.

Overall, the results in this section indicate that short sellers' segmentation increases the predictive ability of aggregate short interest. I find that in countries where short sellers hold more industry-concentrated portfolios, aggregate short interest has a stronger ability to predict market returns. I also provide more direct empirical evidence on the mechanics of the segmentation channel by showing that short interest in the most shorted stocks predicts returns and negative earnings news on other, less shorted stocks. Nevertheless, part of the predictive power of short interest is likely to be driven by short sellers' tendency to avoid trading on their systematic information to avoid bearing systematic risks.

5.2. Other contributors to predictable returns

The previous section demonstrates that short sellers' segmentation plays an important role in a negative relation between aggregate short interest and future market returns. However, this predictability can also arise from other sources. To establish the importance of the segmentation channel, in this section, I test whether these other factors can subsume the predictive ability of aggregate short interest.[22]

5.2.1. Market-wide short sale constraints

First, I examine how limits to arbitrage in the form of market-wide short sale constraints affect the predictive power of aggregate short interest. Boehmer et al. (2018) show that market-wide short sale constraints in the form of exchange regulations, such as naked short sale bans, uptick rules, and centralized equity lending markets, increase the predictive ability of short sales at the stock level. A similar mechanism can work at the market level.

[22] Ideally, I would like to estimate the relative impact of each factor on the predictive power of *SII*. However, variations in these factors have different natures (cross-sectional vs. time-series), which makes such comparison problematic.

Boehmer et al. (2018) show that naked short sale bans, uptick rules, and centralized stock lending markets affect the difficulty of short selling in all securities, thus increasing the informativeness of short interest at the stock level. These regulations have been extensively studied in the literature (e.g., Diether, Lee, and Werner, 2009b; Beber and Pagano, 2013; Huszár and Prado, 2019). Boehmer et al. (2018) describe each short sale regulation in detail and discuss how they differ across countries. The differences in these regulations across countries can generate differences in the predictive power of aggregate short interest.

I run the following pooled global panel regressions:

$$RET_{i,t:t+n} = \alpha + \beta_1 SII_{i,t-1} + \beta_2 Regulation_{i,t-1} + \beta_3 SII_{i,t-1} \times Regulation_{i,t-1} + \beta_4 RET_{i,t-n:t-1} + \varepsilon_{i,t}, \tag{8}$$

where $Regulation_{i,t-1}$ is one of the three short sale regulation dummies—$NakedBan_{i,t-1}$, $Uptick_{i,t-1}$, or $CentralizedLending_{i,t-1}$—that equals one if a naked short sale ban, an uptick rule, or a centralized stock lending market is in place in country *i* for week *t–1*, and zero otherwise. Boehmer et al. (2018) provide a timeline when specific short sale regulations become operational in different countries. If market-wide short sale constraints increase the predictive power of aggregate short interest, I expect to see significant negative coefficients on *SII×NakedBan* and *SII×Uptick* and a significant positive coefficient on *SII×CentralizedLending*.

Table 6 reports the results. In most regression specifications, the interaction coefficients are either statistically insignificant or have "wrong" signs. For example, columns (3) and (6) of Panel A show significantly positive interaction coefficients, indicating that *SII* has a weaker impact on future market returns in countries with naked short sale bans. This result is inconsistent with the expectation that market-wide short sale constraints would increase the predictive power of aggregate short interest. Only 2 out of 18 regression specifications in columns (5)–(6) of Panel C produce significant interaction coefficients consistent with this expectation. Overall, these results indicate that market-wide short sale constraints have a limited ability to explain why short interest predicts market returns and are thus unlikely to subsume the effect of the segmentation channel.

[Table 6 here]

5.2.2. Funding constraints

Funding constraints constitute another type of limits to arbitrage that short sellers face (Richardson et al., 2017). Time-series variation in these constraints can create time-series variation in the predictive power of aggregate short interest. To rule out this possibility, I orthogonalize variations in short interest to the variations in funding constraints and check whether the residual short interest still predicts market returns. If it does, funding constraints cannot fully explain the predictive power of *SII*.[23]

I use two alternative measures of funding constraints that are available internationally. The first measure is the TED spread defined as the difference between the London Interbank Offered Rate (LIBOR) and the risk-free government bond rate. The second measure, the LIBOR-OIS spread, is the difference between LIBOR and an overnight interest swap (OIS) rate.[24] Both measures have been used in the literature to gauge the severity of funding constraints (e.g., Brunnermeier, 2009; Gorton and Metrick, 2012; Rösch, Subrahmanyam, and van Dijk, 2017). In each country, I run the following regression to orthogonalize short interest to funding constraints:

$$Ln(EWSI)_t = \alpha + \beta_1 t + \beta_2 FundConstr_t + \varepsilon_t, \tag{9}$$

[23] I choose to orthogonalize aggregate short interest rather than to control for funding constraints in a multivariate setting because the objective of my analysis is to understand whether existing factors can explain the predictive power of aggregate short interest rather than to compare the effects of short interest to that of funding constraints.

[24] I measure TED and LIBOR-OIS spreads using the interest rates denominated in the local currencies. For all European countries, I use the same benchmark Euro rates. All data come from Datastream.

where $FundConstr_{i,t-1}$ denotes one of the two measures of funding constraints, $TED_{i,t-1}$ or $LIBOR\text{-}OIS_{i,t-1}$. All variables are measured daily. The fitted residual from this regression represents the part of the aggregate short interest that is orthogonal to the time trend, *t*, and variations in funding constraints. I denote this orthogonal part of the aggregate short interest *Orthogonal SII*. For each week, I calculate the average daily *Orthogonal SII* and use it in regressions.

In Table 7, I regress market returns on *Orthogonal SII*. In all regression specifications, short interest orthogonal to variations in funding constraints has a significant negative impact on future returns. The statistical and economic significance of coefficients on *Orthogonal SII* and *EWSI* is similar to the one on the original *SII* and *EWSI* coefficients in Panel B of Table 3. These results indicate that funding constraints absorb very little predictive power of the aggregate short interest.

[Table 7 here]

5.2.3. Time-varying aggregate risk premium

The predictive power of aggregate short interest can stem from its exposure to the time-varying aggregate risk premium. I thus orthogonalize short interest to variations in predictors of the aggregate risk premium and check whether this residual short interest still predicts market returns. I use five popular predictors of market returns for *SII* orthogonalization: dividend yield (*DY*), short-term government bond yield (*Gvt bond yld*), term spread (*Term spread*), global variance risk premium (*VRP*), and the ratio of gold to platinum prices (*GP*). These variables have been shown to be related to the aggregate risk premium (e.g., Ang and Bekaert, 2007; Hjalmarsson, 2010; Rapach et al. , 2013; Bollerslev, Marrone, Xu, and Zhou, 2014; Huang and Kilic, 2019).

I run the following orthogonalization procedure within each country:

$$Ln(EWSI)_t = \alpha + \beta_1 t + \beta_2 Ln(DY)_t + \beta_3 Gvt\ bond\ yld_t + \beta_4 Term\ spread_t + \beta_5 VRP_t + \beta_6 Ln(GP)_t + \varepsilon_t, \tag{10}$$

where all variables are measured daily. I use all five predictors simultaneously because a combination of individual predictors can more successfully capture the risk premium (Rapach, Strauss, and Zhou, 2010). The fitted residual from this regression is the aggregate short interest orthogonal to the variation in five predictors of aggregate risk premium. I compute its average daily value per week and use it in all regressions.

Panel A of Table 8 reports the results of regressing market returns on the orthogonal part of short interest. Significantly negative coefficients on *Orthogonal SII* and *EWSI* at all return horizons suggest that the aggregate short interest continues to predict market returns after excluding the variation in aggregate risk premium. The statistical and economic significance of *Orthogonal SII* is lower compared to the original *SII* in Panel B of Table 3. For example, column (1) indicates that the coefficient on *Orthogonal SII* equals –37.89 (*t*-statistic of –2.69) at a monthly horizon compared to a corresponding coefficient of –60.37 (*t*-statistic of –3.29) for *SII* in column (1) of Panel B, Table 3. At the same time, the statistical and economic significance of *Orthogonal EWSI* in columns (4)–(6) corresponds well to that of the original *EWSI* in columns (4)–(6) of Panel B, Table 3. Overall, these results suggest that the time-varying aggregate risk premium absorbs part of the predictive power of short interest but leaves a significant proportion of it unexplained.

[Table 8 here]

5.2.4. Market sentiment and investor disagreement

Since aggregate investor disagreement and market sentiment predict market returns (e.g., Yu, 2011; Huang, Jiang, Tu, and Zhou, 2015; Jiang, Lee, Martin, and Zhou, 2019), they could potentially subsume the predictive power of aggregate short interest. Similar to the previous section, I orthogonalize short interest to the measures of investor disagreement and sentiment to check whether the remaining variation in short interest still predicts market returns. Following Yu (2011), I measure aggregate investor disagreement every month as the value-weighted average standard deviation of the analyst forecasts of earnings-per-share across all stocks (from I/B/E/S). Market sentiment is measured monthly as the value of the OECD Consumer Confidence Index or the OECD Business Confidence Index. The indexes quantify the expectations of the businesses and households in different countries regarding future economic prospects. These country-specific indexes are similar in spirit to the Michigan Consumer Sentiment Index used by Stambaugh, Yu, and Yuan (2012), Greenwood and Shleifer (2014), and Huang et al. (2015) to measure market sentiment in the United States.

In Table 9, I regress market returns on the orthogonal part of short interest. I find that *Orthogonal SII* continues to predict market returns. Thus, aggregate investor disagreement and market sentiment cannot explain the predictive power of aggregate short interest.

[Table 9 here]

Overall, short sale regulations, funding constraints, time-varying aggregate risk premium, aggregate investor disagreement, and market sentiment do not fully explain the predictive power of aggregate short interest. These findings indicate that market return predictability driven by short interest retains a significant proportion of this predictability unexplained. Given that short sellers' segmentation increases the predictive power of aggregate short interest, I conclude that the segmentation channel is likely to play an important role in this market-level phenomenon

6. Trading on predictable market returns

If predictable market returns partly arise because of mispricing, one should be able to profitably trade on this predictability, conditional on having access to timely short interest data. I create a simple trading signal based on historical short interest data from Markit, which is available in real time, and check whether one can make profits by following this signal. At the beginning of each week, I calculate *EWSI* across stocks in the previous week. *EWSI* is not demeaned as *SII* and therefore does not suffer from a look-ahead bias. If the previous-week *EWSI* is one standard deviation higher than its historical average, it signifies market overpricing.[25] Upon observing this signal, I sell market index futures and hold this position from 1 to 12 weeks.

Fig. 4 plots the average futures returns after observing the trading signal. The number of trading signals during the sample period varies from 9 in the United States and Switzerland to more than 300 in Taiwan. To represent each country in the plot equally, I first calculate the average futures returns within each of the 27 countries and then average these values across all countries. I plot these average market index futures returns in Fig. 4. They are reliably negative at all investment horizons. In an average country, shorting futures following the trading signal brings a return of 0.91% per month and 3.21% per quarter. These results are not driven by a few countries with extreme values. Average futures returns are negative at most horizons in 15 out of 27 countries. I emphasize that the trading strategy does not have a look-ahead bias; anyone who has access to the Markit database can re-create the trading signal that I construct.[26]

[Fig. 4 here]

[25] I require at least 100 weekly historical observations to obtain the signal.

[26] Alternatively, one should be able to construct a similar trading signal using publicly available short-selling data. In 24 out of 32 countries from my sample, exchanges regularly release data on short sales to the public. Delay in publication of these data rarely exceeds one week. Given that aggregate short interest predicts negative futures returns over several months, these delays would unlikely make the strategy unimplementable.

It is hard to estimate the statistical significance of the above results because the trading strategy returns lack counterfactual returns for statistical comparison. Nevertheless, I can compare futures and investment strategy returns after observing the trading signal with similar returns after observing no such signal. I do it in Table A9 of Appendix A and find that futures returns following the trading signal are negative and significantly lower than all other returns (Panel A), while strategy excess returns (returns from selling index futures and buying local government bonds) are positive and significantly higher than all other returns (Panel B).

My findings indicate that investors can profit from trading in index futures by following the signal from past short sales. Importantly, index futures are very liquid instruments that have virtually no short sale constraints, indicating that this strategy is easy to implement. Of course, if short sellers were using this strategy, there would be no return predictability in the first place. The most likely reason why short sellers do not implement this strategy is because it is associated with bearing systematic risks, which they try to avoid (see Section 5.1. for discussion). However, there are other potential reasons why they are not doing it.

For example, they may overlook this opportunity if they are not perfectly rational. Von Beschwitz and Massa (2020) show that short sellers are prone to the disposition effect, suggesting that these sophisticated investors do not always behave rationally. Second, short sellers may have limited capital for investments, and it could be more profitable to use this capital to trade in individual stocks with high market betas as these stocks essentially provide in-built leverage on the systematic information on which short sellers can trade. Finally, short sellers may not have sufficient attention capacity to process already acquired information for trading in market index securities. Theoretical literature suggests that specialized investors with limited attention capacity tend to learn about the stocks that they already hold (e.g., Van Nieuwerburgh and Veldkamp, 2009, 2010). Therefore, even if short sellers uncover systematic information, processing this information for trading in market index instruments requires additional attention resources, which are not necessarily readily available. While beyond the scope of the present paper, it would be interesting to investigate these reasons in future research.

7. Conclusion

Rapach et al. (2016) find that high levels of aggregate short interest predict low market returns in the United States. I explore whether this pattern exists in other markets globally and what can justify its existence. I show that aggregate short interest is negatively associated with future market returns in 30 out of 32 countries, and in 20 countries, this association is statistically significant. On average, a one-standard-deviation increase in *SII* leads to a 0.60% lower market return in the next month. However, this negative relation varies substantially across countries, and this variability depends on the level of short sellers' segmentation. I show that short interest in the most shorted stocks predicts returns on other, less shorted stocks, indicating a slow diffusion of short sellers' systematic information across stock segments. Overall, the short sellers' segmentation appears to be an important channel contributing to the persistence of predictable market returns.

References

Aitken, M.J., Frino, A., McCorry, M.S., Swan, P.L., 1998. Short sales are almost instantaneously bad news: Evidence from the Australian Stock Exchange. Journal of Finance 53, 2205–2223.

Akbas, F., Boehmer, E., Erturk, B., Sorescu, S., 2017. Short interest, returns, and unfavourable fundamental information. Financial Management 46, 455–486.

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5, 31–56.

Ang, A., Bekaert, G., 2007. Stock return predictability: Is it there? Review of Financial Studies 20, 651–707.

Ang, A., Hodrick, R.J., Xing, Y., Zhang, X., 2006. The cross-section of volatility and expected returns. Journal of Finance 61, 259–299.

Ang, A., Hodrick, R.J., Xing, Y., Zhang, X., 2009. High idiosyncratic volatility and low returns: International and further U.S. evidence. Journal of Financial Economics 91, 1–23.

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

Atmas, A., Basak, S., Ruan, F., 2020. Dynamic equilibrium with costly short-selling and lending market. Working paper, Purdue University and London Business School.

Baker, M., Wurgler, J., 2007. Investor sentiment in the stock market. Journal of Economic Perspectives 21, 129–151.

Baker, M., Wurgler, J., Yuan, Y., 2012. Global, local, and contagious investor sentiment. Journal of Financial Economics 104, 272–287.

Beber, A., Pagano, M., 2013. Short-selling bans around the world: Evidence from the 2007–09 crisis. Journal of Finance 68, 343–381.

Beneish, M.D., Lee, C.M.C., Nichols, D.C., 2015. In short supply: Short-sellers and stock returns. Journal of Accounting and Economics 60, 33–57.

von Beschwitz, B., Massa, M., 2020. Biased short: Short sellers' disposition effect and limits to arbitrage. Journal of Financial Markets forthcoming.

Blocher, J., Reed, A.V., van Wesep, E.D., 2013. Connecting two markets: An equilibrium framework for shorts, longs, and stock loans. Journal of Financial Economics 108, 302–322.

Boehmer, E., Huszár, Z.R., Jordan, B.D., 2010. The good news in short interest. Journal of Financial Economics 96, 80–97.

Boehmer, E., Huszár, Z.R., Wang, Y., Zhang, X., 2018. Are shorts equally informed? A Global Perspective. Working paper, Singapore Management University, National University of Singapore, Shanghai University of Finance, and Tsinghua University.

Boehmer, E., Jones, C.M., Zhang, X., 2008. Which shorts are informed? Review of Financial Studies 63, 491–527.

Boehmer, E., Wu., J., 2013. Short selling and price discovery process. Review of Financial Studies 26, 287–322.

Bollerslev, T., Marrone, J., Xu, L., Zhou, H., 2014. Stock return predictability and variance risk premia: Statistical inference and international evidence. Journal of Financial and Quantitative Analysis 49, 633–661.

Brunnermeier, M.K., 2009. Deciphering the liquidity and credit crunch 2007–2008. Journal of Economic Perspectives 23, 77–100.

Choi, N., Fedenia, M., Skiba, H., Sokolyuk, T., 2017. Portfolio concentration and performance of institutional investors worldwide. Journal of Financial Economics 123, 189–208.

Christophe, S.E., Ferri, M.G., Angel, J.J., 2004. Short selling prior to earnings announcements. Journal of Finance 59, 1845–1876.

Christophe, S.E., Ferri, M.G., Hsieh, J., 2010. Informed trading before analyst downgrades: Evidence from short sellers. Journal of Financial Economics 95, 85–106.

Cremers, C.J.M., Petajisto, A., 2009. How active is your fund manager? A new measure that predicts performance. Review of Financial Studies 22, 3329–3365.

D'Avolio, G., 2002. The market for borrowing stock. Journal of Financial Economics 66, 271–306.

Desai, H., Ramesh, K., Thiagarajan, S.R., Balachandran, B.V., 2002. An investigation of the informational role of short interest in the Nasdaq market. Journal of Finance 57, 2263–2287.

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

Diether, K.B., Lee, K.-H., Werner, I.M., 2009a. Short-sale strategies and return predictability. Review of Financial Studies 22, 575–607.

Diether, K.B., Lee, K.-H., Werner, I.M., 2009b. It's SHO time! Short-sale price tests and market quality. Journal of Finance 64, 37–73.

Doshi, H., Elkamhi, R., Simutin, M., 2015. Managerial activeness and mutual fund performance. Review of Asset Pricing Studies 5, 156–184.

Driscoll, J.C., Kraay, A.C., 1998. Consistent covariance matrix estimation with spatially dependent panel data. Review of Economics and Statistics 80, 549–560.

Engelberg, J.E., McLean, D., Pontiff, J., Ringgenberg, M.C., 2019. Are cross-sectional predictors good market-level predictors? Working paper, University of California, San Diego, Georgetown University, Boston College, and University of Utah.

Engelberg, J.E., Reed, A.V., Ringgenberg, M.C., 2018. Short-selling risk. *Journal of Finance* 73, 755–786.

Figlewski, S., 1981. The informational effects of restrictions on short sales: Some empirical evidence. Journal of Financial and Quantitative Analysis 4, 463–476.

Figlewski, S., Webb, G.P., 1993. Options, short sales, and market completeness. Journal of Finance 48, 761–777.

Fong, K.Y.L., Holden, C.W., Trzcinka, C.A., 2017. What are the best liquidity proxies for global research? Review of Finance 21, 1355–1401.

Gargano, A., Sotes-Paladino, J., Verwijmeren, P., 2019. The profits and losses of short sellers. Working paper, University of Melbourne and Erasmus School of Economics.

Gârleanu, N., Pedersen, L.H., 2020. Active and passive investing. Working paper, University of California, Berkley and Copenhagen Business School.

Gastineau, G.L., 2010. The exchange-traded funds manual, vol. 186. John Wiley & Sons, New York.

Glasserman, P., Mamaysky, H., 2019. Investor information choice with macro and micro information. Working paper, Columbia Business School.

Gorton, G., Metrick, A., 2012. Securitized banking and the run on repo. Journal of Financial Economics 104, 425–451.

Greenwood, R., Shleifer, A., 2014. Expectations of returns and expected returns. Review of Financial Studies 27, 714–746.

Griffin, J.M., Kelly, P.J., Nardari, F., 2010. Do market efficiency measures yield correct inferences? A comparison of developed and emerging markets. Review of Financial Studies 23, 3225–3277.

Han, B., Li, G., 2017. Aggregate implied volatility spread and stock market returns. Working paper, University of Toronto.

Harrison, M., Kreps, D., 1978. Speculative investor behavior in a stock market with heterogeneous expectations. Quarterly Journal of Economics 92, 323–336.

Henkel, S.J., Martin, J.S., Nardari, F., 2011. Time-varying short-horizon predictability. Journal of Financial Economics 99, 560–580.

Hirshleifer, D., Lim, S.S., Teoh, S.H., 2009. Driven to distraction: Extraneous events and underreaction to earnings news. Journal of Finance 64, 2289–2325.

Hjalmarsson, E., 2010. Predicting global stock returns. Journal of Financial and Quantitative Analysis 45, 49–80.

Hodrick, R.J., 1992. Dividend yield and expected stock returns: Alternative procedures for inference and measurement. Review of Financial Studies 5, 357–386.

Hong, H., Li, F.W., Ni, S.X., Scheinkman, J., Yan, F., 2016. Days to cover and stock returns. Working paper, Columbia University, Singapore Management University, Hong Kong Baptist University, and Quantitative Investment Strategies.

Hong, H., Torous, W., Valkanov, R., 2007. Do industries lead stock markets? Journal of Financial Economics 83, 367–396.

Hou, K., Karolyi, G.A., Kho, B.-C., 2011. What factors drive global stock returns? Review of Financial Studies 24, 2527–2574.

Huang, D., Jiang, F., Tu, J., Zhou, G., 2015. Investor sentiment aligned: A powerful predictor of stock returns. Review of Financial Studies 28, 791–837.

Huang, D., Kilic, M., 2019. Gold, platinum, and expected stock returns. Journal of Financial Economics 132, 50–75.

Huang, S., O'Hara, M., Zhong, Z., 2020. Innovation and informed trading: Evidence from industry ETFs. Working paper, University of Hong Kong, Cornell University, University of Technology Sydney, and University of Melbourne.

Huszár, Z.R., Prado, M.P., 2019. An analysis of over-the-counter and centralized stock lending markets. Journal of Financial Markets 43, 31–53.

Huszár, Z.R., Tan, R.S.K., Zhang, W., 2017. Do short sellers exploit industry information? Journal Empirical Finance 41, 118–139.

Ince, O., Porter, R.B., 2003. Individual equity return data from Thomson Datastream: Handle with care! Journal of Financial Research 29, 463–479.

Jiang, F., Lee, J., Martin, X., Zhou, G., 2019. Manager sentiment and stock returns. Journal of Financial Economics 132, 126–149.

Jones, C.M., Lamont, O.A., 2002. Short-sale constraints and stock returns. Journal of Financial Economics 66, 207–239.

Kacperczyk, M., Sialm, C., Zheng, L., 2005. On the industry concentration of actively managed equity mutual funds. Journal of Finance 60, 1983–2011.

Karpoff, J.M., Lou, X., 2010. Short sellers and financial misconduct. Journal of Finance 65, 1879–1913.

Kolasinski, A.D., Reed, A.V., Ringgenberg, M.C., 2013. A multiple lender approach to understanding supply and search in the equity lending market. Journal of Finance 68, 559–595.

Kostakis, A., Magdalinos, T., Stamatogiannis, M.P., 2015. Robust econometric inference for stock return predictability. Review of Financial Studies 28, 1506–1553.

Kothari, S.P., Shanken, J., 1997. Book-to-market, dividend yield, and expected market returns: A time-series analysis. Journal of Financial Economics 44, 169–203.

Lo, A.W., MacKinlay, A.C., 1990. When are contrarian profits due to stock market overreaction? Review of Financial Studies 3, 175–205.

Miller, E.M., 1977. Risk, uncertainty, and divergence of opinion. Journal of Finance 32, 1151–1168.

Newey, W.K., West, K.D., 1987. A simple, positive, semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Prado, M.P., Saffi, P.A.C., Sturgess, J., 2016. Ownership structure, limits to arbitrage, and stock returns: Evidence from equity lending markets. Review of Financial Studies 29, 3211–3244.

Priestley, R., 2019. Short interest, macroeconomic variables, and aggregate stock returns. Working paper, Norwegian Business School.

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

Rapach, D.E., Strauss, J.K., Zhou, G., 2010. Out-of-sample equity premium prediction: combination forecasts and links to the real economy. Review of Financial Studies 23, 821–862.

Rapach, D.E., Strauss, J.K., Zhou, G., 2013. International stock return predictability: What is the role of the United States. Journal of Finance 68, 1633–1662.

Richardson, S., Saffi, P.A.C., Sigurdsson, K., 2017. Deleveraging risk. Journal of Financial and Quantitative Analysis 52, 2491–2522.

Rösch, D.M., Subrahmanyam, A., van Dijk, M.A., 2017. The dynamics of market efficiency. Review of Financial Studies 30, 1151–1187.

Saffi, P.A.C., Sigurddson, K., 2011. Price efficiency and short selling. Review of Financial Studies 24, 821–852.

Scheinkman, J., Xiong, W., 2003. Overconfidence and speculative bubbles. Journal of Political Economy 111, 1183–1219.

Stambaugh, R.F., 1999. Predictive regressions. Journal of Financial Economics 54, 375–421.

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

Sutcliffe, C.M.S., 2006. Stock index futures, third edition. Ashgate Publishing, Hampshire.

Torous, W., Valkanov, R., Yan, S., 2004. On predicting stock returns with nearly integrated explanatory variables. Journal of Business 77, 937–966.

Van Nieuwerburgh, S., Veldkamp, L., 2009. Information immobility and the home bias puzzle. Journal of Finance 64, 1187–1215.

Van Nieuwerburgh, S., Veldkamp, L., 2010. Information acquisition and portfolio under-diversification. Review of Economic Studies 77, 779–805.

Wang, X., Yan, X.S., Zheng, L., 2020. Shorting flows, public disclosure, and market efficiency. Journal of Financial Economics 135, 191–212.

Yu, J., 2011. Disagreement and return predictability of stock portfolios. Journal of Financial Economics 99, 162–183.

Table 1

Data coverage and statistics on aggregate short-selling activity

The table provides statistics on the data coverage and the estimates of the aggregate short-selling activity for the period from July 1, 2006 to December 31, 2016. *MSCI group* denotes developed (DM) and emerging (EM) markets, as classified by Morgan Stanley Capital International (MSCI). *N Stocks* is the total number of unique stocks in the sample. *N Weeks* is the total number of weekly observations in the sample. *EWSI* is the equal-weighted average daily short interest (as a percentage of shares outstanding) across all stocks. The last three columns denote the availability of data on market index futures, options, and ETFs in Datastream for a given country during the sample period. Table A3 of Appendix A provides precise availability periods for each instrument.

Country	*MSCI group*	*N Stocks*	*N Weeks*	*EWSI*	Availability of market index instruments		
					Futures	Options	ETFs
Australia	DM	979	535	1.04%	Yes	Yes	Yes
Austria	DM	62	532	1.24%	Yes	No	No
Belgium	DM	116	542	0.65%	Yes	No	No
Brazil	EM	141	512	0.08%	Yes	Yes	Yes
Canada	DM	1,066	537	1.93%	Yes	No	Yes
Denmark	DM	136	514	0.65%	Yes	Yes	No
Finland	DM	124	533	1.27%	No	No	No
France	DM	564	542	1.05%	Yes	Yes	Yes
Germany	DM	551	533	1.17%	Yes	No	No
Greece	EM	57	507	0.04%	Yes	Yes	No
Hong Kong	DM	894	523	0.51%	Yes	No	Yes
Ireland	DM	46	537	0.24%	No	No	No
Israel	DM	131	496	0.09%	No	No	No
Italy	DM	302	533	0.88%	Yes	Yes	No
Japan	DM	4,057	524	0.62%	Yes	Yes	Yes
Korea	EM	1,273	521	0.38%	Yes	Yes	Yes
Mexico	EM	98	537	0.39%	Yes	No	Yes
Netherlands	DM	133	542	1.50%	Yes	Yes	No
New Zealand	DM	88	528	0.33%	Yes	No	No
Norway	DM	221	522	0.86%	Yes	No	Yes
Poland	EM	120	467	0.14%	Yes	Yes	No
Portugal	DM	36	542	0.60%	Yes	No	No
Russia	EM	69	526	0.01%	Yes	Yes	No
Singapore	DM	386	537	0.35%	Yes	Yes	No
South Africa	EM	206	534	0.49%	Yes	Yes	Yes
Spain	DM	151	538	1.06%	Yes	Yes	Yes
Sweden	DM	363	533	0.86%	Yes	Yes	Yes
Switzerland	DM	285	532	1.12%	Yes	Yes	Yes
Taiwan	EM	963	513	0.32%	Yes	Yes	Yes
Turkey	EM	191	529	0.24%	Yes	No	No
United Kingdom	DM	1,406	535	1.06%	Yes	No	No
United States	DM	6,494	546	3.55%	Yes	Yes	Yes

Table 2

Determinants of short sellers' segmentation

Panel A reports the OLS estimates from the following regressions within each country:

$$Overweight_{i,t} = \alpha + \beta_1 Lendable_{i,t-1} + \beta_2 LenderConc_{i,t-1} + \beta_3 ShortRisk_{i,t-1} + \beta_4 FHT_{i,t-1} + \beta_5 IVol_{i,t-1} + \beta_6 BM_{i,t-1} + \beta_7 LogMCap_{i,t-1} + FE_i + FE_t + \varepsilon_{i,t},$$

where $Overweight_{i,t}$ is the difference between value weights of industry *i* in the aggregate short sellers' portfolio and in the market portfolio at the end of week *t*, $w^s_{i,t} - w^m_{i,t}$. *Lendable* denotes the value-weighted average number of stocks available for lending (standardized by the number of shares outstanding) for an average for stocks in industry *i* at the end of week *t–1*. *LenderConc* is the value-weighted average weekly metric of lenders concentration from Markit for stocks in industry *i* in week *t–1*. *ShortRisk* is the value-weighted average yearly lending fee variability for stocks in industry *i* up to week *t–1*. *FHT* is the value-weighted average monthly illiquidity measure from Fong et al., (2017) for stocks in industry *i* up to week *t–1*. *IVol* is is the value-weighted average monthly idiosyncratic volatility estimated from the model of Hou et al. (2011) for stocks in industry *i* up to week *t–1*. *BM* is the value-weighted average last year-end book-to-market ratio for stocks in industry *i* available at the end of week *t–1*. *LogMCap* is the cumulative market capitalization of all stocks in industry *i* at the end of week *t–1*. All regressions include week fixed effects. All independent variables are standardized to have a mean of zero and a standard deviation of one within each country. The sample period is from July 2006 to December 2016. I cluster standard errors at the industry level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panels B and C report results for the similar pooled global panel regressions that impose the restriction of homogeneous slope coefficients across countries. All regressions include country and week fixed effects. I cluster standard errors at the country-industry level. Panel B employs a full sample, while Panel C excludes the United States from the sample.

Panel A: Country-level regressions

Country	*Lendable*	*LenderConc*	*ShortRisk*	*FHT*	*IVol*	*BM*	*LogMCap*	NObs	Adj. R^2
Australia	0.23**	–0.21***	–0.01	–0.17	0.19**	–0.09	–0.43	44,522	2.53%
Austria	1.75***	–0.20	0.56*	0.42	–0.24	–0.19	–0.11	11,405	11.11%
Belgium	2.12*	–1.09**	–0.96	0.04	0.15	–2.10	–0.98	17,514	14.68%
Brazil	1.94***	0.98***	–0.00	–0.23	0.72	–0.95	–0.60	6,794	15.86%
Canada	0.09	–0.06*	–0.00	0.14	–0.21	–0.00	0.06	46,747	0.94%
Denmark	2.88**	–0.48*	0.21*	–1.03**	1.08*	0.06	–2.92	13,780	13.55%
Finland	0.50	–0.23	–0.12	0.08	0.43	–0.13	–0.02	20,675	0.43%
France	0.44***	–0.09***	0.02	0.03	–0.06	0.06	–0.35*	48,450	3.99%
Germany	0.31***	–0.08**	0.01	0.00	–0.01	–0.01	–0.18	40,444	2.21%
Greece	2.94	2.59***	0.33	–0.25	0.56	0.65	–2.19	5,133	3.92%
Hong Kong	0.50**	–0.25***	0.01	0.02	0.24	–0.12	–0.46	42,468	4.18%
Ireland	3.48*	–1.82**	–0.97	–0.09	–0.21	–0.23	–1.41	6,417	3.24%
Israel	1.36	–0.49	0.59***	–0.49*	0.91	0.11	–1.42	11,796	3.07%
Italy	0.50**	–0.27***	–0.07	–0.02	–0.03	–0.07	–0.39	30,931	0.86%
Japan	0.12**	–0.11***	0.00	–0.03*	0.08***	–0.01	–0.17**	67,077	4.54%
Korea	0.07	–0.23**	0.06	0.04	0.46**	–0.03	–0.13	42,618	2.64%

Table 2
Continued

Country	*Lendable*	*LenderConc*	*ShortRisk*	*FHT*	*IVol*	*BM*	*LogMCap*	NObs	Adj. R^2
Mexico	0.99**	–1.31***	–0.12	–0.03	0.98*	0.02	–1.27*	14,976	14.68%
Netherlands	1.93***	–0.84***	0.08	0.12	0.60**	–0.12	–1.20**	19,116	21.39%
New Zealand	4.13***	–0.58**	–0.20	0.84	–0.61	–0.62	–1.47	13,125	27.52%
Norway	1.10**	–0.22	0.07	0.15	1.16	0.02	–0.10	16,692	3.66%
Poland	2.35**	–0.82**	0.14	0.04	0.22	–0.21	–1.00	9,622	9.07%
Portugal	5.45***	–1.12**	0.39	–1.13*	0.81*	0.57	–3.95***	6,822	15.07%
Russia	4.46	4.69**	–0.35	3.33	0.08	–5.00	–2.53	3,597	6.93%
Singapore	0.29	–0.49***	0.12	–0.12	0.04	0.01	–0.40	29,131	1.44%
South Africa	0.26	–0.55***	–0.03	–0.30	0.50***	–0.21*	0.03	22,620	1.99%
Spain	1.37***	–0.26*	0.08	0.11	–0.09	0.08	–0.75	20,824	8.12%
Sweden	1.16***	–0.37***	–0.04	–0.22	0.16	–0.29**	–1.45***	28,488	13.27%
Switzerland	0.37***	–0.40***	–0.08	–0.09	0.17	–0.13	–0.65**	29,565	5.80%
Taiwan	0.22*	–0.10	0.04	0.00	0.26	0.08	0.13	29,217	3.33%
Turkey	0.25	–0.14	–0.05	0.04	–0.26	0.51	0.78	20,944	4.36%
United Kingdom	0.13**	–0.22***	0.00	–0.11**	0.13**	0.02	–0.33*	52,734	1.50%
United States	0.04***	–0.04***	–0.00	–0.01***	0.01	–0.00	–0.09**	304,758	5.24%
Median estimate	0.75***	–0.24***	0.00	–0.01	0.17	–0.02	–0.45	33,719	7.22%
(*t*-stat)	(2.21)	(–2.34)	(0.48)	(–0.02)	(1.07)	(–0.44)	(–1.08)		

Panel B: Pooled global panel regressions

Independent variable	*Lendable*	*LenderConc*	*ShortRisk*	*FHT*	*IVol*	*BM*	*LogMCap*	NObs	Adj. R^2
Estimate	0.54***	–0.15**	–0.02	–0.00	0.20***	–0.03	–0.30**	1,086,323	2.90%
(*t*-stat)	(2.77)	(–2.08)	(–0.61)	(–0.12)	(3.40)	(–1.38)	(–2.40)		

Panel C: Pooled global panel regressions, excluding the United States

Independent variable	*Lendable*	*LenderConc*	*ShortRisk*	*FHT*	*IVol*	*BM*	*LogMCap*	NObs	Adj. R^2
Estimate	0.77***	–0.23***	–0.01	–0.01	0.23***	–0.04	–0.47***	781,565	3.72%
(*t*-stat)	(4.29)	(–3.02)	(–0.33)	(–0.37)	(3.82)	(–1.59)	(–3.93)		

Table 3

Predictive regressions of market returns on aggregate short interest

Panel A reports the OLS estimates of β_1 from the following time-series regressions within each country:

$$RET_{t:t+n} = \alpha + \beta_1 SII_{t-1} + \beta_2 RET_{t-n:t-1} + \beta_3 US\ RET_{t-n:t-1} + \varepsilon_t\ ,$$

where $RET_{t:t+n}$ is the cumulative market index return in country *i* for *n* weeks starting from week *t*, SII_{t-1} is the average daily value of the short interest index defined as the fitted residual in Eq. (1) for week *t–1* (see Section 3.2), $RET_{t-n:t-1}$ is the cumulative market index return for the past *n* weeks ending in week *t–1*, and $US\ RET_{t-n:t-1}$ is the cumulative S&P 500 index return for the last *n* weeks ending in week *t–1*. In columns (4)–(6), I run similar regressions but replace *SII* with its non-detrended version, *EWSI*, defined as in Table 1. *SII* and *EWSI* are standardized to have a mean of zero and a standard deviation of one. The regressions for the United States exclude $US\ RET_{t-n:t-1}$. All coefficient estimates are in basis points. The sample period is from July 2006 to December 2016. I use heteroskedasticity- and autocorrelation-robust Newey-West standard errors with 12 lags. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel B reports results for the similar pooled global panel regressions that impose the restriction of homogeneous slope coefficients across countries and omits $US\ RET_{t-n:t-1}$. Figures in parentheses (brackets) are *t*-statistics based on heteroskedasticity- and autocorrelation-robust Newey-West standard errors with 12 lags (based on Driscoll-Kraay standard errors that are robust to cross-country correlations, heteroskedasticity, and autocorrelation).

Panel A: Time-series regressions

	Independent variable: SII_{t-1}			Independent variable: $EWSI_{t-1}$		
	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)
Expected *SII* (*EWSI*) sign	–	–	–	–	–	–
# negative *SII*s (*EWSI*s)	29	30	30	29	30	30
# negative & significant	13	21	20	10	16	16
# positive *SII*s (*EWSI*s)	3	2	2	3	2	2
# positive & significant	1	1	1	1	0	0
Coefficients on *SII* (*EWSI*)						
Australia	–57.02	–125.48*	–199.03*	–47.36	–106.00	–176.49*
Austria	–71.62*	–182.56**	–273.81**	–70.49*	–181.04**	–289.46**
Belgium	–111.19***	–200.64***	–205.10**	–98.91***	–177.12***	–189.35**
Brazil	34.39	64.35	73.97	66.75*	98.68	139.02
Canada	–103.50**	–219.29***	–374.87***	–37.08	–51.05	–89.71
Denmark	–61.01	–169.60***	–283.21***	–44.48	–141.18**	–250.52***
Finland	8.25	–38.69	–128.81	16.85	–11.49	–97.00
France	–11.33	–53.86	–83.73	–33.78	–86.56	–125.13
Germany	–64.09*	–153.87***	–259.50***	–69.07*	–148.79***	–247.62***
Greece	–195.70***	–370.38**	–519.84**	–233.90***	–457.19***	–572.04***
Hong Kong	–78.50	–149.70	–223.83	–89.28*	–156.03*	–232.34
Ireland	–144.25**	–318.12**	–463.63***	–203.49***	–432.95***	–606.77***
Israel	–31.59	–85.32	–89.43	–49.07	–116.71*	–140.43
Italy	–60.65	–173.24**	–214.56**	–54.40	–157.20**	–193.91**
Japan	–97.10*	–203.42**	–297.27**	–80.89	–168.76	–248.72*
Korea	–68.63	–142.16*	–172.91	–53.08	–99.08	–110.13
Mexico	–46.87	–107.06*	–165.43**	–44.10	–95.06	–122.43
Netherlands	–20.56	–99.67	–130.60	–29.40	–116.79	–162.61
New Zealand	–61.22**	–147.54**	–215.96***	–85.33***	–213.39***	–309.27***
Norway	–55.46	–158.24**	–254.19**	–45.81	–148.98	–260.70*
Poland	78.34*	180.73***	295.92**	15.93	17.29	13.71

Table 3
Continued

	Independent variable: SII_{t-1}			Independent variable: $EWSI_{t-1}$		
	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)
Coefficients on *SII* (*EWSI*)						
Portugal	–105.53**	–237.25***	–315.70***	–94.27**	–208.79**	–285.90**
Russia	–172.28***	–376.19***	–592.12***	–130.46**	–296.69**	–482.21**
Singapore	–63.33	–140.16	–221.45	–47.19	–126.15	–224.09
South Africa	–50.36	–117.07*	–193.16**	–33.08	–69.30	–115.95
Spain	–71.39*	–183.58***	–173.82*	–48.40	–129.17**	–117.38
Sweden	–18.69	–97.03*	–151.25*	–20.09	–101.55**	–160.63**
Switzerland	–58.54*	–154.10***	–221.82***	–58.44	–143.26**	–207.54**
Taiwan	–30.79	–37.43	–77.13	–7.33	–3.16	11.02
Turkey	–5.95	–16.41	–25.90	–13.09	–45.31	–93.48
United Kingdom	–40.78	–119.10	–165.07	–45.80	–120.57	–167.83
United States	–100.45**	–238.32***	–341.51***	–101.86**	–229.15***	–323.41***

Panel B: Pooled global panel regressions

	Independent variable: SII_{t-1}			Independent variable: $EWSI_{t-1}$		
	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)
SII or *EWSI*	–60.37***	–140.02***	–203.24***	–58.40***	–136.99***	–198.78***
	(–7.39)	(–8.99)	(–9.15)	(–6.92)	(–8.49)	(–8.72)
	[–3.29]	[–4.43]	[–5.42]	[–3.75]	[–5.02]	[–6.17]
NObs	16,752	16,624	16,498	16,843	16,715	16,589
Adj. R^2	1.07%	2.59%	4.81%	1.01%	2.50%	4.75%

Table 4

Predictive power of aggregate short interest and short sellers' segmentation

The table reports the OLS estimates of β_1 and β_3 from the following global panel regressions:

$$RET_{i,t:t+n} = \alpha + \beta_1 SII_{i,t-1} + \beta_2 Segmented_{i,t-1} + \beta_3 SII \times Segmented_{i,t-1} + \beta_4 RET_{i,t-n:t-1} + \varepsilon_{i,t}\,,$$

where $RET_{i,t:t+n}$, $RET_{i,t-n:t-1}$, and $SII_{i,t-1}$ are defined as in Table 3. $Segmented_{i,t-1}$ is a dummy variable that equals one (zero) if short sellers in country *i* are characterized by above-median (below-median) industry segmentation in week *t-1* across my sample countries. The industry segmentation is measured by *IndustryConc* or *IndustrySpec* defined as in Section 3.3. Panel A (Panel B) reports regression results in which short sellers' segmentation is measured by *IndustryConc* (*IndustrySpec*). In columns (4)–(6), I run similar regressions but replace *SII* with its non-detrended version, *EWSI*, defined as in Table 1. *SII* and *EWSI* are standardized to have a mean of zero and a standard deviation of one within each country. All coefficient estimates are in basis points. The sample period is from July 2006 to December 2016. Figures in brackets are *t*-statistics based on Driscoll-Kraay standard errors that are robust to cross-country correlations, heteroskedasticity, and autocorrelation. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: *IndustryConc* as a measure of short sellers' segmentation						
	Independent variable: SII_{t-1}			Independent variable: $EWSI_{t-1}$		
	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)
SII or *EWSI*	–46.43**	–111.34***	–175.96***	–42.09**	–99.25***	–160.13***
	[–2.12]	[–3.00]	[–3.88]	[–2.37]	[–3.54]	[–4.90]
SII or *EWSI* × *Segmented*	–28.23**	–58.54***	–57.56*	–34.13**	–79.26***	–84.67***
	[–2.21]	[–2.69]	[–1.87]	[–2.56]	[–3.37]	[–2.72]
NObs	16,738	16,610	16,484	16,754	16,626	16,500
Adj. R^2	1.15%	2.74%	4.92%	1.11%	2.75%	4.92%
Panel B: *IndustrySpec* as a measure of short sellers' segmentation						
	Independent variable: SII_{t-1}			Independent variable: $EWSI_{t-1}$		
	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)
SII or *EWSI*	–46.59**	–116.75***	–185.13***	–41.37**	–101.34***	–161.11***
	[–2.15]	[–3.20]	[–4.15]	[–2.28]	[–3.43]	[–4.71]
SII or *EWSI* × *Segmented*	–28.43**	–49.18**	–40.80	–36.18**	–76.69***	–84.26***
	[–2.06]	[–2.08]	[–1.27]	[–2.56]	[–3.14]	[–2.71]
NObs	16,738	16,610	16,484	16,754	16,626	16,500
Adj. R^2	1.15%	2.73%	4.90%	1.12%	2.76%	4.93%

Table 5
Cross-predictability effects in stocks with different levels of short interest

The table reports the OLS estimates of β_1 and β_3 from the following global panel regressions:

$$TercX\ RET_{i,t:t+n} = \alpha + \beta_1 Terc1\ EWSI_{i,t-1} + \beta_2 Terc2\ EWSI_{i,t-1} + \beta_3 Terc3\ EWSI_{i,t-1} + \beta_4 Terc1\ RET_{i,t-n:t-1} + \beta_5 Terc2\ RET_{i,t-n:t-1} + \beta_6 Terc3\ RET_{i,t-n:t-1} + \varepsilon_{i,t}\,,$$

where *Terc1 EWSI*, *Terc2 EWSI*, and *Terc3 EWSI* are weekly *EWSI*s (defined in Table 1) in stocks with the lowest, medium, and highest short interests, respectively. $TercX\ RET_{t:t+n}$ denotes the equal-weighted return for n weeks starting from week t on the portfolio of stocks categorized into tercile X based on their weekly short interest. *TercX EWSI*s are standardized to have a mean of zero and a standard deviation of one within each country. Greece and Russia are omitted from the sample due to the lack of variation in *Terc1 EWSI*. All coefficient estimates are in basis points. The sample period is from July 2006 to December 2016. Figures in brackets are t-statistics based on Driscoll-Kraay standard errors that are robust to cross-country correlations, heteroskedasticity, and autocorrelation. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively

	Dependent variable: *Terc1* portfolio returns			Dependent variable: *Terc2* portfolio returns			Dependent variable: *Terc3* portfolio returns		
	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Terc1 EWSI	14.47	14.48	27.02	15.29	7.08	3.67	8.25	3.91	–1.00
	[1.01]	[0.52]	[0.82]	[0.86]	[0.21]	[0.09]	[0.40]	[0.10]	[–0.03]
Terc2 EWSI	–33.60*	–60.86**	–85.51**	–24.43	–41.46	–56.56	–26.46	–53.64	–64.49
	[–1.88]	[–2.19]	[–2.34]	[–1.19]	[–1.25]	[–1.28]	[–1.12]	[–1.55]	[–1.54]
Terc3 EWSI	–34.80***	–92.77***	–153.03***	–53.54***	–115.83***	–179.45***	–55.04***	–120.68***	–178.59***
	[–3.01]	[–3.59]	[–4.60]	[–3.74]	[–3.80]	[–4.58]	[–3.36]	[–3.69]	[–4.18]
NObs	12,924	10,454	8,397	12,921	10,451	8,420	12,913	10,411	8,341
Adj. R^2	8.65%	9.57%	16.45%	5.13%	8.01%	16.12%	3.00%	6.14%	14.48%

Table 6

Predictive power of aggregate short interest and market-wide short sale constraints

The table reports the OLS estimates of β_1 and β_3 from the following global panel regressions:

$$RET_{i,t:t+n} = \alpha + \beta_1 SII_{i,t-1} + \beta_2 Regulation_{i,t-1} + \beta_3 SII \times Regulation_{i,t-1} + \beta_4 RET_{i,t-n:t-1} + \varepsilon_{i,t}\,,$$

where $RET_{i,t:t+n}$, $RET_{i,t-n:t-1}$, and $SII_{i,t-1}$ are defined as in Table 3. $Regulation_{i,t-1}$ is one of the short sale regulation dummies—$NakedBan_{i,t-1}$, $Uptick_{i,t-1}$, or $CentralizedLending_{i,t-1}$—that equals one if a naked short sale ban, an uptick rule, or a centralized stock lending market is in place in country *i* in week *t–1*, and zero otherwise. In columns (4)–(6), I run similar regressions but replace *SII* with its non-detrended version, *EWSI*, defined as in Table 1. *SII* and *EWSI* are standardized to have a mean of zero and a standard deviation of one within each country. All coefficient estimates are in basis points. The sample period is from July 2006 to December 2016. Figures in brackets are *t*-statistics based on Driscoll-Kraay standard errors that are robust to cross-country correlations, heteroskedasticity, and autocorrelation. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Naked short sale bans

	Independent variable: SII_{t-1}			Independent variable: $EWSI_{t-1}$		
	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)
SII or *EWSI*	-60.22^{***}	-150.22^{***}	-236.30^{***}	-57.80^{***}	-150.37^{***}	-241.19^{***}
	[–2.85]	[–4.18]	[–5.84]	[–2.82]	[–4.19]	[–6.06]
SII or *EWSI* ×	–0.08	18.44	59.37^{**}	–0.78	24.81	77.75^{**}
NakedBan	[–0.00]	[0.75]	[2.23]	[–0.05]	[0.88]	[2.43]
NObs	16,752	16,624	16,498	16,843	16,715	16,589
Adj. R^2	1.09%	2.63%	4.92%	1.03%	2.56%	4.92%

Panel B: Uptick rules

	Independent variable: SII_{t-1}			Independent variable: $EWSI_{t-1}$		
	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)
SII or *EWSI*	-51.08^{**}	-125.45^{***}	-184.81^{***}	-50.37^{**}	-124.93^{***}	-186.77^{***}
	[–2.49]	[–3.71]	[–4.69]	[–2.55]	[–3.71]	(–4.72)
SII or *EWSI* ×	–32.76	–51.51	–65.26	–29.78	–44.71	–44.09
Uptick	[–1.55]	[–1.51]	[–1.48]	[–1.06]	[–0.92]	[–0.69]
NObs	16,752	16,624	16,498	16,843	16,715	16,589
Adj. R^2	1.14%	2.69%	4.92%	1.06%	2.56%	4.79%

Panel C: Centralized stock lending markets

	Independent variable: SII_{t-1}			Independent variable: $EWSI_{t-1}$		
	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)
SII or *EWSI*	-62.50^{***}	-144.59^{***}	-209.56^{***}	-66.86^{***}	-157.19^{***}	-228.05^{***}
	[–3.26]	[–4.41]	[–5.39]	[–4.10]	[–5.37]	[–6.38]
SII or *EWSI* ×	4.83	10.48	14.61	19.15	45.58^{**}	65.91^{**}
CentralizedLending	[0.37]	[0.48]	[0.49]	[1.37]	[1.98]	[2.08]
NObs	16,752	16,624	16,498	16,843	16,715	16,589
Adj. R^2	1.07%	2.60%	4.82%	1.03%	2.57%	4.84%

Table 7

Predictive power of aggregate short interest orthogonal to variations in funding constraints

The table reports the OLS estimates of β_1 from the following global panel regressions:

$$RET_{i,t:t+n} = \alpha + \beta_1 Orthogonal\ SII_{i,t-1} + \beta_2 RET_{i,t-n:t-1} + \varepsilon_{i,t}\,,$$

where $RET_{i,t:t+n}$ and $RET_{i,t-n:t-1}$ are defined as in Table 3. *Orthogonal SII* is the short interest index orthogonal to the variation in one of the two measures of funding constraints: TED or LIBOR-OIS spread. TED spread is the difference between the LIBOR and the risk-free government bond rate in the respective country. LIBOR-OIS spread is the difference between the LIBOR and the overnight interest swap (OIS) rate. Daily *Orthogonal SII* in a country is the fitted residual, $\hat{\varepsilon}_t$, from the following OLS regression: $Ln(EWSI)_t = \alpha + \beta_1 \times t + \beta_2 FundConstr_t + \varepsilon_t$, where $t = 1, \ldots, T$ and signifies the day of the sample period. $FundConstr_t$ denotes the value of the TED or LIBOR-OIS spread on day t. $Orthogonal\ SII_{i,t-1}$ is the average daily *Orthogonal SII* in country i for week $t-1$. In columns (4)–(6), I run similar regressions but replace *Orthogonal SII* with its non-detrended version, *Orthogonal EWSI* (*EWSI* is defined in Table 1). *Orthogonal SII* and *EWSI* are standardized to have a mean of zero and a standard deviation of one within each country. All coefficient estimates are in basis points. The sample period is from July 2006 to December 2016. In Panel A, the sample consists of 28 countries that have valid data on TED spread in Datastream. In Panel B, the sample consists of 25 countries that have valid data on LIBOR-OIS spread in Datastream. Figures in brackets are *t*-statistics based on Driscoll-Kraay standard errors that are robust to cross-country correlations, heteroskedasticity, and autocorrelation. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Short interest orthogonal to variations in TED spread

	Independent variable: $Orthogonal\ SII_{t-1}$			Independent variable: $Orthogonal\ EWSI_{t-1}$		
	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)
Orthogonal SII or *EWSI*	-58.37^{***}	-136.35^{***}	-199.35^{***}	-60.95^{***}	-142.04^{***}	-209.74^{***}
	[–3.36]	[–4.36]	[–5.12]	[–3.59]	[–4.58]	[–5.50]
NObs	14,569	14,461	14,355	14,650	14,542	14,436
Adj. R^2	0.97%	2.43%	4.71%	1.06%	2.65%	5.15%

Panel B: Short interest orthogonal to variations in LIBOR-OIS spread

	Independent variable: $Orthogonal\ SII_{t-1}$			Independent variable: $Orthogonal\ EWSI_{t-1}$		
	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)
Orthogonal SII or *EWSI*	-59.59^{***}	-139.19^{***}	-201.89^{***}	-54.17^{***}	-126.88^{***}	-186.69^{***}
	[–3.05]	[–4.14]	[–5.03]	[–3.31]	[–4.29]	[–5.03]
NObs	12,018	11,926	11,834	12,099	12,007	11,915
Adj. R^2	0.99%	2.58%	5.02%	0.83%	2.17%	4.65%

Table 8

Predictive power of aggregate short interest orthogonal to variations in time-varying aggregate risk premium

The table reports the OLS estimates of β_1 from the following global panel regressions:

$$RET_{i,t:t+n} = \alpha + \beta_1 Orthogonal\ SII_{i,t-1} + \beta_2 RET_{i,t-n:t-1} + \varepsilon_{i,t}\,,$$

where $RET_{i,t:t+n}$ and $RET_{i,t-n:t-1}$ are defined as in Table 3. *Orthogonal SII* is the short interest index orthogonal to the variation in five popular predictors of time-varying aggregate risk premium: *DY*, *Gvt bond yld*, *Term spread*, *VRP*, and *GP*. *DY* is the dividend yield on the market index. *Gvt bond yld* is the yield on the three-month government bond. *Term spread* is the difference between yields of the ten-year and three-month government bonds. *VRP* is the global variance risk premium defined as in Bollerslev et al. (2014). *GP* is the ratio of gold to platinum prices. Daily *Orthogonal SII* in a country is the fitted residual, $\hat{\varepsilon}_t$, from the following OLS regression: $Ln(EWSI)_t = \alpha + \beta_1 \times t + \beta_2 Ln(DY)_t + \beta_3 Gvt\ bond\ yld_t + \beta_4 Term\ spread_t + \beta_5 VRP_t + \beta_6 Ln(GP)_t + \varepsilon_t$, where $t = 1, \ldots, T$ and signifies the day of the sample period. $Orthogonal\ SII_{i,t-1}$ is the average daily *Orthogonal SII* in country i for week $t-1$. In columns (4)–(6), I run similar regressions but replace *Orthogonal SII* with its non-detrended version, *Orthogonal EWSI* (*EWSI* is defined in Table 1). *Orthogonal SII* and *EWSI* are standardized to have a mean of zero and a standard deviation of one within each country. All coefficient estimates are in basis points. The sample period is from July 2006 to December 2016. The sample consists of 22 countries that have valid data on the predictors of time-varying aggregate risk premium in Datastream. Figures in brackets are t-statistics based on Driscoll-Kraay standard errors that are robust to cross-country correlations, heteroskedasticity, and autocorrelation. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	Independent variable: $Orthogonal\ SII_{t-1}$			Independent variable: $Orthogonal\ EWSI_{t-1}$		
	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)
Orthogonal SII or *EWSI*	-37.89^{***}	-90.38^{***}	-127.79^{***}	-55.96^{***}	-132.70^{***}	-189.50^{***}
	[–2.96]	[–5.10]	[–5.30]	[–3.68]	[–5.34]	[–5.76]
NObs	11,299	11,215	11,133	11,362	11,278	11,196
Adj. R^2	0.45%	1.21%	2.36%	0.95%	2.58%	4.25%

Table 9

Predictive power of aggregate short interest orthogonal to variations in investor disagreement and sentiment

The table reports the OLS estimates of β_1 from the following global panel regressions:

$$RET_{i,t:t+n} = \alpha + \beta_1 Orthogonal\ SII_{i,t-1} + \beta_2 RET_{i,t-n:t-1} + \varepsilon_{i,t}\,,$$

where $RET_{i,t:t+n}$ and $RET_{i,t-n:t-1}$ are defined as in Table 3. *Orthogonal SII* is the short interest index orthogonal to the variation in the measure of aggregate dispersion of opinion or market sentiment. Aggregate investor disagreement is measured monthly as the value-weighted average standard deviation of the analyst forecasts of earnings-per-share across all stocks (see Yu, 2011). Market sentiment is measured monthly as the value of the OECD Consumer Confidence Index or the OECD Business Confidence Index. The indexes quantify the expectations of the businesses and households regarding future economic prospects. *Orthogonal SII* in a country is the fitted residual, $\hat{\varepsilon}_t$, from the following OLS regression: $Ln(EWSI)_t = \alpha + \beta_1 \times t + \beta_2 Predictor_t + \varepsilon_t$, where $t = 1, \ldots, T$ and signifies the month of the sample period. $EWSI_t$ is the weekly *EWSI* (defined as in Table 2) at the end of month t. $Predictor_t$ denotes the value of the predictor at the end of month t. Panels A, B, and C report results for short interest orthogonalized to variations in aggregate dispersion of opinion, Consumer Confidence Index, and Business Confidence Index, respectively. In columns (4)–(6), I run similar regressions but replace *Orthogonal SII* with its non-detrended version, *Orthogonal EWSI*. *Orthogonal SII* and *EWSI* are standardized to have a mean of zero and a standard deviation of one within each country. All coefficient estimates are in basis points. The sample period is from July 2006 to December 2016. In Panels A, the sample consists of 26 countries that have valid data on analyst forecasts in I/B/E/S. In Panels B and C, the sample consists of 26 and 27 countries that have valid data on the Consumer Confidence Index and Business Confidence Index in OECD Data, respectively. Figures in brackets are t-statistics based on Driscoll-Kraay standard errors that are robust to cross-country correlations, heteroskedasticity, and autocorrelation. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Short interest orthogonal to variations in aggregate dispersion of opinion

	Independent variable: *Orthogonal* SII_{t-1}			Independent variable: *Orthogonal* $EWSI_{t-1}$		
	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)
Orthogonal SII or *EWSI*	–53.30**	–139.48***	–204.46***	–53.05**	–128.82***	–196.14***
	[–2.38]	[–3.21]	[–3.44]	[–2.29]	[–2.82]	[–3.30]
NObs	2,067	2,048	2,029	2,067	2,048	2,029
Adj. R^2	1.77%	3.04%	5.56%	1.76%	2.64%	5.27%

Panel B: Short interest orthogonal to variations in market sentiment proxied by Consumer Confidence Index

	Independent variable: *Orthogonal* SII_{t-1}			Independent variable: *Orthogonal* $EWSI_{t-1}$		
	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)
Orthogonal SII or *EWSI*	–65.43***	–148.06***	–201.52***	–54.43***	–129.24***	–186.76***
	[–3.21]	[–3.39]	[–3.71]	[–3.07]	[–4.06]	[–4.78]
NObs	2,991	2,965	2,939	3,007	2,981	2,955
Adj. R^2	2.07%	2.88%	5.19%	1.73%	2.25%	4.78%

Panel C: Short interest orthogonal to variations in market sentiment proxied by Business Confidence Index

	Independent variable: *Orthogonal* SII_{t-1}			Independent variable: *Orthogonal* $EWSI_{t-1}$		
	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)
Orthogonal SII or *EWSI*	–66.16***	–146.64***	–196.61***	–59.22***	–133.24***	–186.55***
	[–2.86]	[–3.08]	[–3.34]	[–3.21]	[–4.02]	[–4.71]
NObs	2,991	2,965	2,939	3,172	3,145	3,118
Adj. R^2	2.09%	2.81%	5.00%	1.95%	2.41%	4.69%

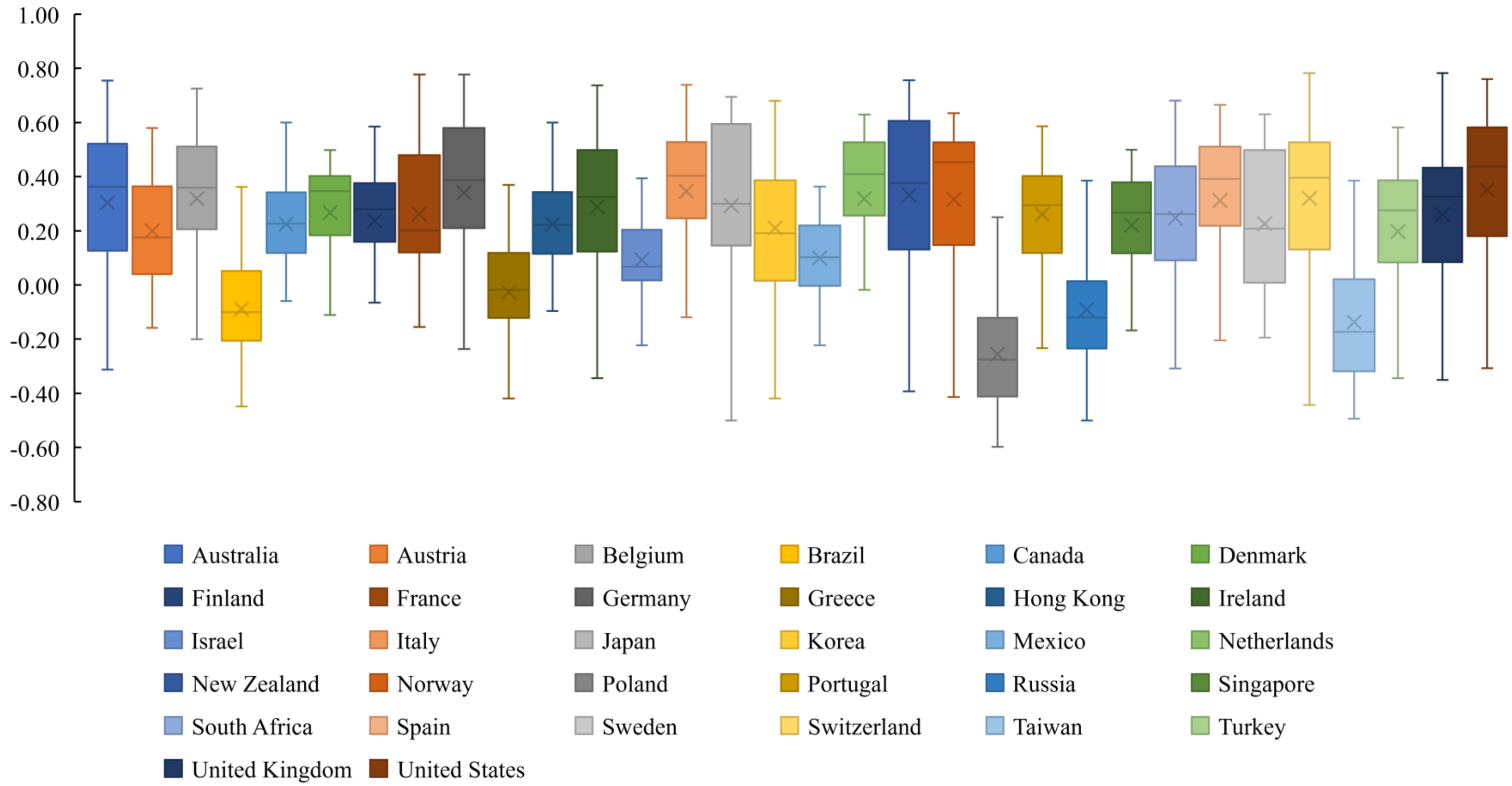


Fig. 1. Cross-country correlations between aggregate short interest indexes (*SII*s)

The figure plots cross-country correlations between daily country-specific *SII*s, defined as the fitted residuals from the OLS regression: $Ln(EWSI)_t = \alpha + \beta \times t + \varepsilon_t$, where t is the day of the sample period and *EWSI* is the equal-weighted short interest across stocks defined as in Table 1. *SII*s are standardized within each country to have a mean of zero and a standard deviation of one. Each box and whisker plot shows the range of correlations between a given country's *SII* and *SII*s in other countries. A cross (a horizontal line) in each box indicates the mean (median) correlation between the given country's *SII* and *SII*s in other countries. The plot is based on the full sample of 32 countries. The sample period for correlation calculations is from July 2006 to December 2016.

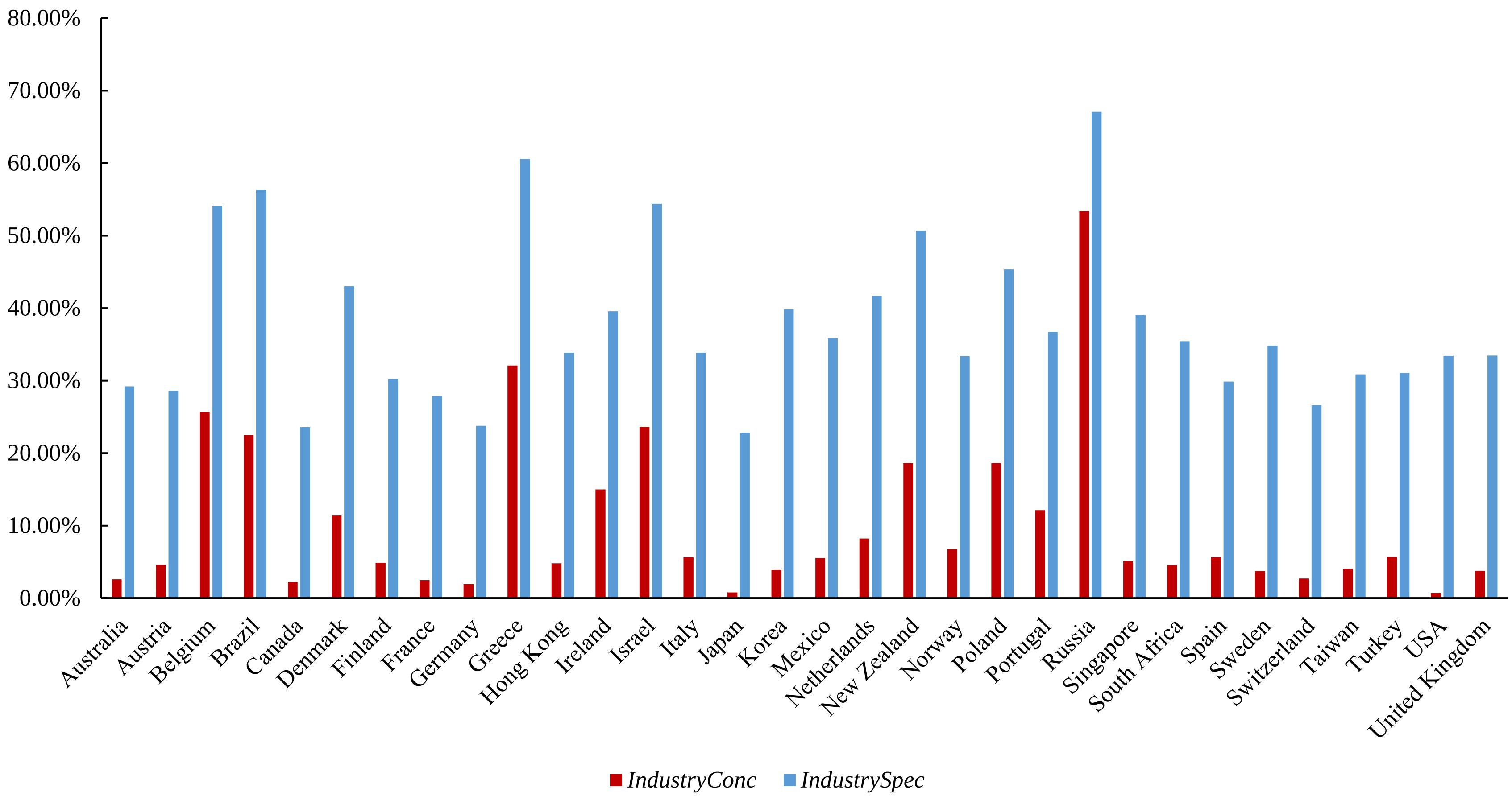
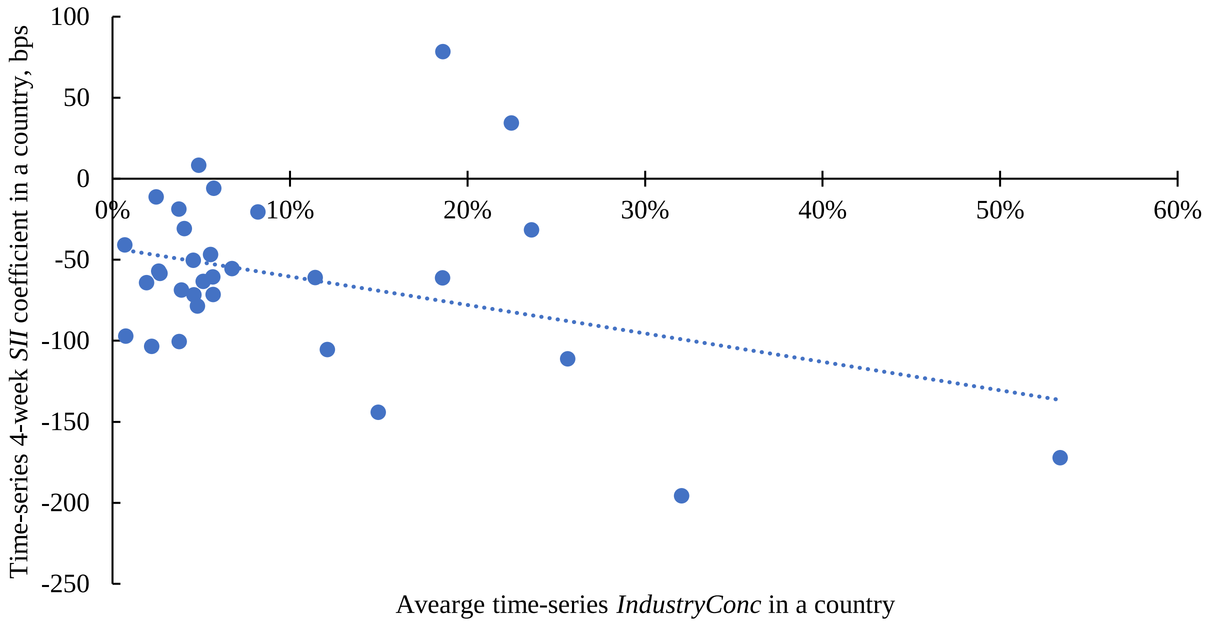


Fig. 2. Short sellers' segmentation

The figure depicts the average short sellers' industry segmentation, or specialization, in 32 countries from July 2006 to December 2016. The segmentation is gauged on a weekly basis by two alternative measures, *IndustryConc* and *IndustrySpec*. $IndustryConc_t = \sum_{i=1}^{n}\left(w_{i,t}^{s} - w_{i,t}^{m}\right)^2$ and $IndustrySpec_t = \frac{1}{2}\sum_{i=1}^{n}\left|w_{i,t}^{s} - w_{i,t}^{m}\right|$, where $w_{i,t}^{s}$, is the value weight of industry *i* in the portfolio consisting of all short positions in a country at the end of week *t* and $w_{i,t}^{m}$ is the value weight of industry *i* in the market portfolio of a country at the end of week *t*. The figure depicts the average weekly values of *IndustryConc* and *IndustrySpec* over the sample period (from July 1, 2006 to December 31, 2016) for each country. *IndustryConc* increases as the collective portfolio of all short sellers becomes more concentrated on a few industries. *IndustrySpec* increases as short sellers start to bet more heavily on particular industries.

Panel A: *IndustryConc* as a measure of short sellers' segmentation and short interest coefficients across countries



Panel B: *IndustrySpec* as a measure of short sellers' segmentation and short interest coefficients across countries

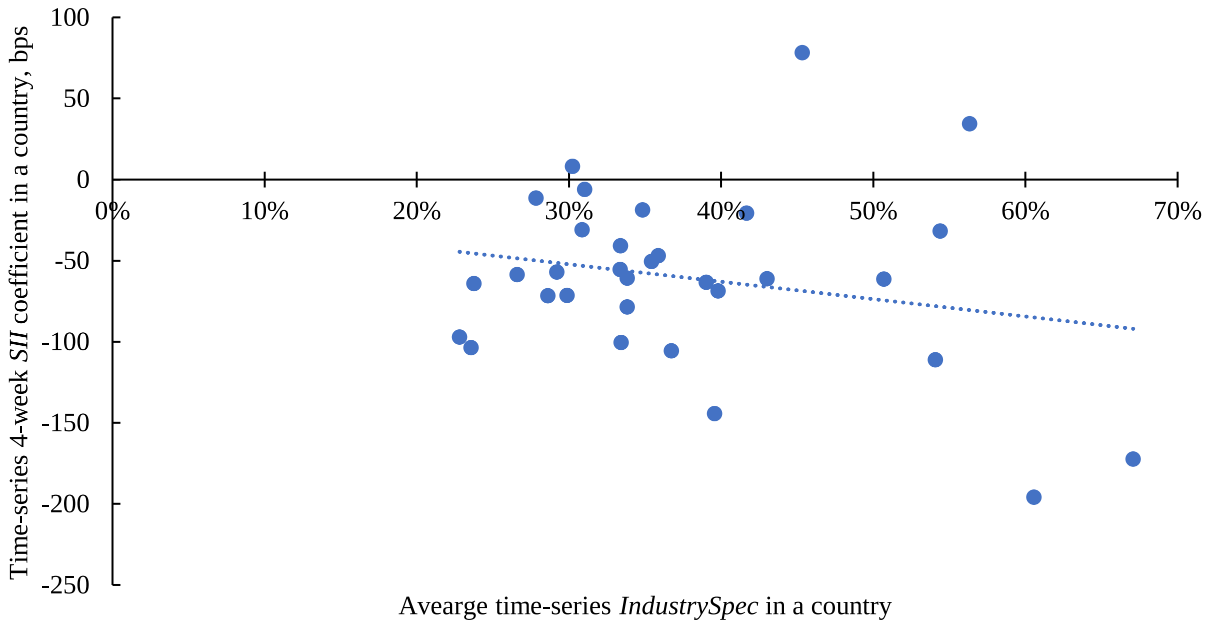


Fig. 3. Short sellers' segmentation and aggregate short interest index (*SII*) coefficients across countries

The figure depicts the scatterplot of time-series four-week *SII* coefficients in a country from Table 3 (vertical axis) against the average time-series short sellers' segmentation measure, *IndustryConc* and *IndustrySpec*, in a country from Fig. 2 (horizontal axis). *SII* is defined as in Table 3; *IndustryConc* and *IndustrySpec* are defined as in Fig. 2.

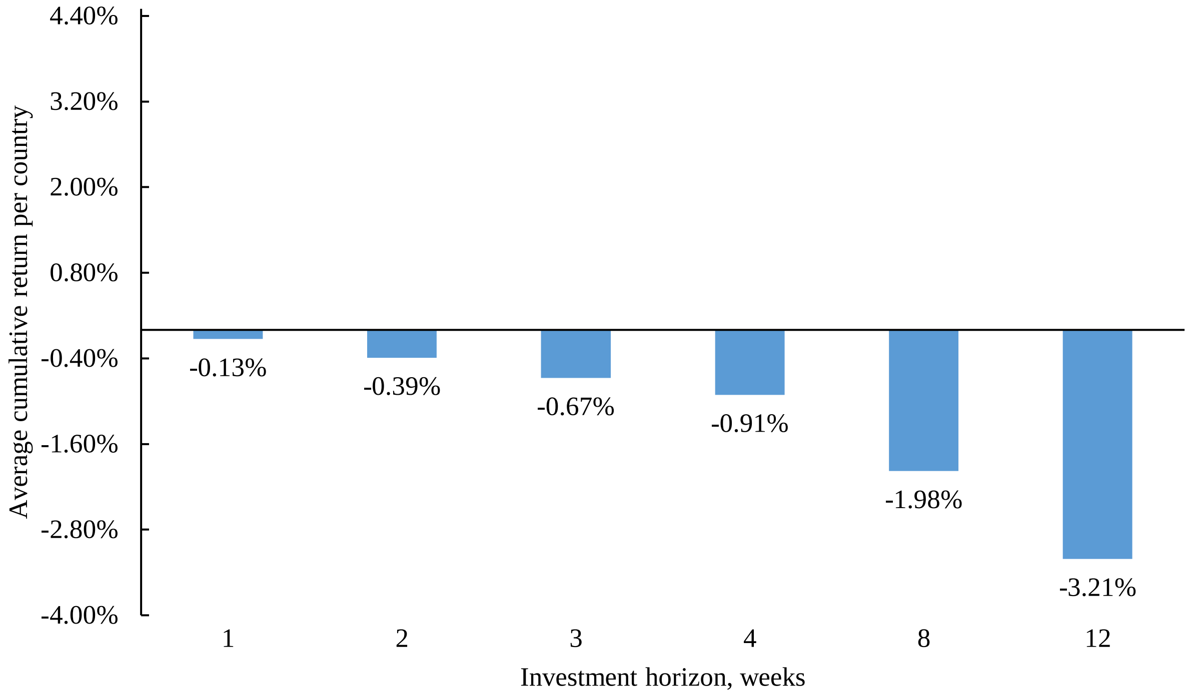


Fig. 4. Market index futures returns following the trading signals from past levels of aggregate short interest

The figure plots the average cumulative returns on market index futures across countries for different investment horizons, after observing a trading signal. The trading signal signifies market overpricing and is obtained when the equal-weighted average short interest (*EWSI*) across stocks in the previous week is at least one standard deviation higher than its average historical value. I require at least 100 weekly historical observations to obtain the signal. Calculations are based on a sample of 27 countries with valid Datastream futures data for at least four consecutive years (Table A3 of Appendix A provides precise availability periods for futures data in each country). The sample period is from July 2006 to December 2016.

Appendix A

Table A1
Market indexes used in the paper
This table reports the country market indexes used in the paper and their corresponding Datastream codes.

Country	Market index	Datastream code
Australia	S&P/ASX 200	ASX200I
Austria	ATX	ATXINDX
Belgium	BEL 20	BGBEL20
Brazil	IBOVESPA	BRBOVES
Canada	TSX Composite	TTOCOMP
Denmark	OMX Copenhagen 20	DKKFXIN
Finland	OMH Helsinki 25	HEXINDX
France	CAC 40	FRCAC40
Germany	DAX 30	DAXINDX
Greece	Athex	GRAGENL
Hong Kong	Hang Seng	HNGKNGI
Ireland	ISEQ	ISEQUIT
Israel	FTSE Israel	WIISRLL
Italy	FTSE MIB	FTSEMIB
Japan	TOPIX	TOKYOSE
Korea	KOSPI 200	KOR200I
Mexico	MEXICO IPC	MXIPC35
Netherlands	AEX	AMSTEOE
New Zealand	NZX 50	NZ50CAP
Norway	OSEAX	OSLOASH
Poland	WIG	POLWIGI
Portugal	PSI 20	POPSI20
Russia	RTSI	RSRTSIN
Singapore	Straits Times	SNGPORI
South Africa	FTSE/JSE All Share	JSEOVER
Spain	IBEX 35	IBEX35I
Sweden	OMX Stockholm 30	SWEDOMX
Switzerland	SMI	SWISSMI
Taiwan	TAIEX	TAIWGHT
Turkey	BIST National 100	TRKISTB
United Kingdom	FTSE 100	FTSE100
United States	S&P 500	S&PCOMP

Table A2

Determinants of short sellers' segmentation

Panel A reports the OLS estimates from the following regressions within each country:

$Overweight_{i,t} = \alpha + \beta_1 Lendable_{i,t-1} + \beta_2 LenderConc_{i,t-1} + \beta_3 ShortRisk_{i,t-1} + \beta_4 FHT_{i,t-1} + \beta_5 Disagr_{i,t-1} + \beta_6 IVol_{i,t-1} + \beta_7 BM_{i,t-1} + \beta_8 LogMCap_{i,t-1} + FE_i + FE_t + \varepsilon_{i,t}$,

where $Disagr_{i,t-1}$ measures investor disagreement in industry i as the value-weighted average standard deviation of the analyst forecasts of earnings-per-share across all stocks in the industry in the month preceeding week $t-1$. I exclude Turkey from the analysis as it has no data on earnings forecasts. All other variables are defined as in Table 2. All regressions include week fixed effects. All independent variables are standardized to have a mean of zero and a standard deviation of one within each country. The sample period is from July 2006 to December 2016. I cluster standard errors at the industry level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panels B and C report results for the similar pooled global panel regressions that impose the restriction of homogeneous slope coefficients across countries. All regressions include country and week fixed effects. I cluster standard errors at the country-industry level. Panel B employs a full sample, while Panel C excludes the United States from the sample.

Panel A: Country-level regressions										
Country	*Lendable*	*LenderConc*	*ShortRisk*	*FHT*	*Disagr*	*IVol*	*BM*	*LogMCap*	NObs	Adj. R^2
Australia	0.26^{**}	-0.35^{***}	–0.14	–0.33	0.21	0.30^{**}	–0.18	–0.57	34,978	4.50%
Austria	1.89^{**}	–1.29	1.17^{***}	1.00^{**}	0.24	–0.84	–1.71	–0.01	4,514	12.50%
Belgium	1.37	-4.75^{**}	-3.71^{*}	–1.10	0.44	–0.58	-17.90^{***}	–0.08	2,880	20.72%
Brazil	2.11^{**}	1.29^{***}	–0.77	0.12	–0.78	0.89	–1.87	–0.91	2,896	15.45%
Canada	0.67	–0.40	0.36	2.46	–0.01	-1.43^{*}	0.07	0.71	9,339	5.60%
Denmark	1.78	-7.11^{***}	–4.88	0.19	0.00	4.22^{*}	1.92	-10.68^{***}	3,708	38.21%
Finland	0.71	-1.25^{*}	–0.40	0.99	0.19	1.61^{**}	–0.43	0.38	9,548	1.14%
France	0.54^{***}	-0.43^{***}	0.36^{**}	0.07	–0.00	–0.18	0.13	-0.61^{**}	24,291	6.74%
Germany	0.32^{***}	–0.09	0.01	–0.02	0.03	–0.03	–0.05	–0.09	27,012	1.20%
Greece	1.34	3.65	-17.83^{**}	4.61	–5.74	1.88	–259.36	-43.00^{**}	540	40.46%
Hong Kong	0.71	-0.98^{***}	0.02	0.29	-0.16^{**}	0.57	–0.05	–0.66	14,372	4.98%
Ireland	2.82	-5.91^{**}	15.45^{**}	2.96	6.05^{**}	–13.68	–31.19	5.88	410	24.73%
Israel	3.31	-3.56^{**}	–1.94	–12.48	30.45	–0.17	2.83	0.99	140	33.42%
Italy	0.61^{*}	-2.20^{**}	8.44	–0.26	0.23	–0.56	–0.26	–0.92	7,603	-0.13%
Japan	0.15^{*}	-0.35^{***}	0.00	–0.04	0.02	0.21^{**}	–0.01	-0.39^{*}	26,406	10.23%
Korea	0.11	–0.64	0.28^{**}	0.26^{*}	0.27	2.02^{***}	0.39	–1.32	10,818	17.13%
Mexico	1.04	-1.60^{***}	1.75	1.14	–0.19	1.82^{**}	0.12	–1.23	3,162	29.23%
Netherlands	1.06^{**}	-2.55^{***}	–0.24	0.61	0.07	0.86	–0.22	0.98	5,827	5.58%
New Zealand	6.28^{***}	–0.46	0.46	1.19	0.51	–1.81	–0.21	–1.14	5,302	36.09%
Norway	1.85	-2.39^{**}	0.92	1.32	0.42	3.70^{*}	1.79	–0.31	4,982	9.85%
Poland	5.20^{*}	-4.23^{**}	–22.59	1.65	–0.51	2.92	0.21	0.31	1,800	16.97%

Table A2
Continued

Country	*Lendable*	*LenderConc*	*ShortRisk*	*FHT*	*Disagr*	*IVol*	*BM*	*LogMCap*	NObs	Adj. R^2
Portugal	3.95***	–14.03**	1.37	0.39	5.12**	4.68*	10.59**	–8.14	1,900	18.22%
Russia	–6.77**	11.68**	–1.29	–6.90*	–0.76	–21.55***	–1.02	4.35	642	10.14%
Singapore	0.05	–2.48***	1.25***	0.88	0.20	0.58	0.57	–1.48	8,703	8.06%
South Africa	–0.30	–2.19***	0.66	–0.70	–0.44*	0.99*	–1.09*	0.31	7,655	9.51%
Spain	1.67***	–1.73**	0.76**	0.93	0.20	–0.95*	0.32	–0.94	6,934	5.33%
Sweden	1.34**	–1.39***	0.04	–0.37	0.12	0.24	–0.61**	–2.43***	13,381	18.49%
Switzerland	0.61**	–1.36***	–0.74	–0.55**	–0.04	0.41	–0.14	–1.41**	9,986	7.56%
Taiwan	0.33	–0.85	0.01	–0.11	0.16	1.49**	0.05	1.65	8,050	10.12%
Turkey										
United Kingdom	0.14	–0.69***	–0.10	–0.18	–0.01	0.57**	–0.07	–0.50	15,931	3.90%
United States	0.04***	–0.10***	–0.01	–0.02	0.01	0.06**	–0.01	–0.15**	174,902	9.03%
Median estimate	0.71**	–1.29***	0.01	0.19	0.07	0.41	–0.05	–0.50	14,471	14.03%
(*t*-stat)	(1.76)	(–2.74)	(0.12)	(0.13)	(0.87)	(1.06)	(–0.37)	(–0.82)		

Panel B: Pooled global panel regressions

Independent	*Lendable*	*LenderConc*	*ShortRisk*	*FHT*	*Disagr*	*IVol*	*BM*	*LogMCap*	NObs	Adj. R^2
Estimate	0.53**	–0.47**	–0.03	0.11	0.05	0.37***	–0.11*	–0.19	451,230	8.99%
(*t*-stat)	(2.25)	(–2.70)	(–0.45)	(0.97)	(1.24)	(3.12)	(–1.70)	(–1.44)		

Panel C: Pooled global panel regressions, excluding the United States

Independent	*Lendable*	*LenderConc*	*ShortRisk*	*FHT*	*Disagr*	*IVol*	*BM*	*LogMCap*	NObs	Adj. R^2
Estimate	0.79***	–0.66***	–0.02	0.18	0.07	0.41***	–0.15*	–0.46**	276,328	9.53%
(*t*-stat)	(2.98)	(–3.51)	(–0.23)	(1.35)	(1.30)	(3.05)	(–1.95)	(–2.37)		

Table A3

Availability of market index instruments in different countries

This table reports the period of data availability on market index futures, options, and ETFs for different countries in my sample. The availability period for futures indicates the period for which continuous time series data on market index futures' returns from Datastream is available for a particular country. The availability period for options indicates the period for which continuous time series data on market index options' implied volatilities from Datastream is available. The availability period for ETFs indicates the period for which short-selling data on market index ETFs from the Markit database is available. I consider an ETF to be a "market index ETF" if it has more than 90% correlation with the primary local market index (see the list of these indexes in Table A1). "Full" availability period indicates that the data is available for the entire sample period from July 1, 2006 to December 31, 2016. "NA" indicates that the data for this country is not available for the entire sample period.

Country	Availability period for the market index instruments		
	Futures	Options	ETFs
Australia	Full	From February 23, 2010	Full
Austria	From February 10, 2014	NA	NA
Belgium	Full	NA	NA
Brazil	Full	From May 23, 2011	From July 12, 2007
Canada	Full	NA	Full
Denmark	From December 12, 2011	From September 11, 2012	NA
Finland	NA	NA	NA
France	Full	July 20, 2007 – June 6, 2010	Full
Germany	Full	NA	NA
Greece	Till June 26, 2015	From July 9, 2013	NA
Hong Kong	Full	NA	Full
Ireland	NA	NA	NA
Israel	NA	NA	NA
Italy	Full	From April 24, 2007	NA
Japan	Full	From March 21, 2012	Full
Korea	Full	From December 1, 2012	Full
Mexico	Full	NA	From August 2, 2013
Netherlands	Full	From September 19, 2013	NA
New Zealand	From June 16, 2014	NA	NA
Norway	Full	NA	From December 27, 2006
Poland	Till June 20, 2014	From July 9, 2013	NA
Portugal	Full	NA	NA
Russia	Full	From April 5, 2011	NA
Singapore	Full	From May 3, 2011	NA
South Africa	Full	From February 23, 2011	Full
Spain	Full	From May 24, 2007	From July 24, 2006
Sweden	Full	From May 24, 2007	Full
Switzerland	Full	Full	Full
Taiwan	Full	From February 26, 2010	Full
Turkey	Full	NA	NA
United Kingdom	Full	NA	NA
United States	Full	Full	Full

Table A4

Predictive regressions of market returns on aggregate short interest: the Wald test by Kostakis et al. (2015)

The table reports the OLS estimates of β_1 from the following time-series regressions within each country:

$$RET_{t:t+n} = \alpha + \beta_1 \widehat{SII}_{t-1} + \beta_2 RET_{t-n:t-1} + \beta_3 US\ RET_{t-n:t-1} + \varepsilon_t,$$

where $RET_{t:t+n}$, $RET_{t-n:t-1}$, and $US\ RET_{t-n:t-1}$ are defined as in Table 3. $\widehat{SII}_{t-1}$ is an instrumental variable for SII_{t-1} (defined in Table 2), whose degree of persistence is explicitly controlled via the IVX estimation procedure (see Kostakis et al., 2015). In columns (4)–(6), I run similar regressions but replace $\widehat{SII}$ with its non-detrended version, $\widehat{EWSI}$ ($EWSI$ is defined as in Table 1). $\widehat{SII}$ and $\widehat{EWSI}$ are standardized to have a mean of zero and a standard deviation of one. The regressions for the United States exclude $US\ RET_{t-n:t-1}$. All coefficient estimates are in basis points. The sample period is from July 2006 to December 2016. Statistical inferences are based on the Wald statistic, which tests the individual significance of each regressor. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Time-series regressions

	Independent variable: $\widehat{SII}_{t-1}$			Independent variable: $\widehat{EWSI}_{t-1}$		
	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)
Expected $\widehat{SII}$ ($\widehat{EWSI}$) sign	–	–	–	–	–	–
# negative $\widehat{SII}$s ($\widehat{EWSI}$s)	30	30	29	30	29	27
# negative & significant	13	21	19	11	17	19
# positive $\widehat{SII}$s ($\widehat{EWSI}$s)	2	2	3	2	3	5
# positive & significant	0	1	1	0	0	0
Coefficients on $\widehat{SII}$ ($\widehat{EWSI}$)						
Australia	–56.95*	–147.54***	–201.15***	–55.77	–146.26***	–207.39***
Austria	–108.17	–245.34**	–433.54***	–110.70	–284.39**	–486.21***
Belgium	–97.16	–71.78	36.79	–91.81*	–70.06	24.11
Brazil	25.54	–29.13	–232.72	57.85	192.44	166.40
Canada	–120.36**	–362.91***	–562.37***	–61.48	–135.18	–262.72
Denmark	–104.70**	–295.17***	–600.68***	–97.60**	–287.70***	–577.39***
Finland	–33.22	–210.08**	–370.93***	–26.01	–211.48	–389.43***
France	–22.08	–85.22	–128.37	–34.56	–104.57	–174.26
Germany	–82.75*	–242.45***	–467.51***	–80.62*	–237.13***	–480.18***
Greece	–201.81**	–320.20***	–348.46**	–274.01***	–373.40***	–447.27***
Hong Kong	–68.77	–121.11	–187.51	–80.10	–140.50	–219.97
Ireland	–154.45***	–338.38***	–622.16***	–221.40***	–442.38***	–806.69***
Israel	–77.15	–70.59	41.27	–61.23	–66.72	12.10
Italy	–80.79	–151.71*	–138.02	–76.03	–140.63	–127.20
Japan	–106.51**	–229.97***	–451.68***	–98.17**	–228.02***	–463.87***
Korea	–88.97**	–112.57**	–131.07*	–57.57	–59.91	–72.61
Mexico	–48.87	–110.53	–180.70***	–44.09	–95.68	–129.43
Netherlands	–46.67	–187.10***	–139.69	–54.32	–205.28***	–175.74*
New Zealand	–73.63**	–176.55***	–252.58***	–96.55***	–208.76***	–273.08***
Norway	–67.26	–245.42***	–287.24***	–63.17	–286.71***	–347.92***
Poland	105.59	298.85***	604.42***	18.00	2.94	61.52
Portugal	–131.99**	–287.05***	–366.45**	–115.00**	–274.77***	–350.58***
Russia	–208.38***	–447.72***	–640.59***	–184.26***	–372.65***	–551.17***
Singapore	–67.32	–208.52	–232.66	–58.13	–205.86***	–224.94**
South Africa	–43.74	–140.54***	–213.85***	–32.43	–91.82	–165.93**

Table A4
Continued

	Independent variable: $\widehat{SII}_{t-1}$			Independent variable: $\widehat{EWSI}_{t-1}$		
	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)
Coefficients on $\widehat{SII}$ ($\widehat{EWSI}$)						
Spain	–97.41	–110.03	–149.28	–87.99	–87.52	–115.06
Sweden	–75.13	–178.46**	–454.66***	–78.12	–203.37***	–521.22***
Switzerland	–85.89***	–169.33***	–230.22***	–89.29**	–174.29***	–238.94***
Taiwan	–43.06	–60.82	–84.09	–6.11	13.83	89.39
Turkey	–12.60	39.18	–164.22	–23.69	–73.95	–167.81
United Kingdom	–54.20	–137.33*	–126.63	–55.85	–128.26***	–135.28**
United States	–121.21***	–263.03***	–396.86***	–109.79***	–218.47***	–322.51***

Table A5

Predictive regressions of market returns on aggregate short interest: Stambaugh (1999) bias

Panel A reports the OLS estimates of β_1 and their estimated Stambaugh (1999) bias from the following time-series regressions within each country:

$$RET_{t:t+n} = \alpha + \beta_1 SII_{t-1} + \varepsilon_t \,,$$

where $RET_{t:t+n}$ and SII_{t-1} are defined as in Table 3. Columns (1)–(3) report the *SII* coefficients and columns (4)–(6) report the estimated Stambaugh (1999) bias for these coefficients. *SII* is standardized to have a mean of zero and a standard deviation of one. The estimation of the Stambaugh (1999) bias follows the procedure outlined in Appendix B. All coefficient estimates are in basis points. The sample period is from July 2006 to December 2016. I use heteroskedasticity- and autocorrelation-robust Newey-West standard errors with 12 lags. Statistical inferences are based on *p*-values that compare *t*-statistics of the actual *SII* coefficients in columns (1)–(3) with the percentiles of *t*-statistics obtained in simulations under the assumption of no predictability ($\beta_1 = 0$). *, **, and *** denote statistical significance according to these *p*-values at the 10%, 5%, and 1% level, respectively. The upper part of Panel A shows the number of countries for which the actual *SII* coefficients and their corresponding *t*-statistics are lower (more negative) than at least 90% of the simulated *SII* coefficients and *t*-statistics obtained under the assumption of no predictability ($\beta_1 = 0$). I run 10,000 simulations per country. Appendix B provides details of the simulations.

Panel B reports the results for the similar pooled global panel regressions that impose the restriction of homogeneous slope coefficients across countries. Figures in parentheses are *p*-values based on the comparison of the *t*-statistics.

Panel A: Time-series regressions

Summary of the time-series simulation results	Return horizon		
	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$
# countries whose actual *SII* coefficients are:			
lower (more negative) than the simulated *SII* coefficients	24	27	29
higher than the simulated *SII* coefficients	8	5	3
# countries whose actual Newey-West *t*-statistics for *SII* coefficients			
lower (more negative) than the simulated Newey-West *t*-statistics	20	22	22
higher than the simulated Newey-West *t*-statistics	12	10	10

	Coefficients on SII_{t-1}			Estimated Stambaugh (1999) bias		
	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)
Country-level *SII (SII* bias*)*						
Australia	–54.57*	–127.18*	–211.50**	0.02	–0.80	–0.96
Austria	–74.84**	–182.34**	–284.91**	0.58	0.96	1.47
Belgium	–107.84***	–197.29***	–232.76***	0.20	1.57	2.02
Brazil	12.46	17.10	–0.62	0.62	0.36	1.05
Canada	–98.54**	–220.4**	–350.41***	0.21	0.19	–0.43
Denmark	–65.07*	–170.32**	–292.74***	0.46	0.04	–0.33
Finland	9.44	–36.89	–126.20	–0.66	–0.80	–1.23
France	–10.51	–44.25	–84.54	0.52	0.93	0.53
Germany	–67.68**	–152.47***	–272.47***	–0.05	1.24	1.48
Greece	–175.65***	–334.27**	–504.61***	2.07	3.97	6.80
Hong Kong	–78.99*	–140.99*	–216.29*	5.71	8.27	10.78
Ireland	–154.97**	–347.02***	–565.77***	–0.84	–1.12	–1.62
Israel	–25.21	–79.76*	–96.92	0.07	–0.12	0.70
Italy	–59.87*	–160.33**	–213.5**	0.67	0.84	1.64
Japan	–93.20*	–202.29**	–308.9**	0.15	0.67	0.09
Korea	–63.56*	–133.15*	–182.32*	0.61	1.61	1.75

Table A5
Continued

	Coefficients on SII_{t-1}			Estimated Stambaugh (1999) bias		
	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)
Country-level *SII (SII* bias*)*						
Mexico	–49.09*	–108.02**	–199.90**	–0.81	0.18	–0.25
Netherlands	–23.29	–88.32	–144.44	0.17	0.77	–0.31
New Zealand	–66.45**	–140.29**	–225.08***	0.63	1.16	1.13
Norway	–58.86*	–151.72**	–237.46**	–0.34	–2.40	–2.49
Poland	85.09	197.75	315.12	1.23	0.70	1.44
Portugal	–104.22**	–235.73***	–347.67***	0.93	1.82	1.71
Russia	–199.83***	–420.76***	–602.84***	–0.23	–1.80	–1.83
Singapore	–73.51	–138.26	–233.02*	2.51	3.82	3.38
South Africa	–47.15	–108.50*	–182.00**	–2.21	–3.75	–3.10
Spain	–66.6**	–154.33**	–176.00**	0.93	0.43	0.10
Sweden	–14.13	–62.19	–101.30	–1.31	–1.22	–1.14
Switzerland	–53.24*	–136.15**	–227.09***	0.74	0.76	0.36
Taiwan	–43.77	–95.95	–139.55	0.56	-0.15	0.14
Turkey	–3.93	0.40	–35.53	0.79	1.39	2.13
United Kingdom	–33.81	–80.67	–142.95	0.99	2.54	2.87
United States	–92.34**	–202.49***	–331.33***	2.40	3.23	2.77

Panel B: Pooled global panel regressions

	Coefficients on SII_{t-1}			Estimated Stambaugh (1999) bias		
	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)
SII (*SII* bias)	–61.53***	–139.66***	–218.64***	–0.25	–0.61	–0.86
	(0.00)	(0.00)	(0.00)			
NObs (NSim)	16,752	16,624	16,498	10,000	10,000	10,000
Adj. R^2	1.04%	2.59%	4.16%			

Table A6
Predictive power of aggregate short interest in recessions

The table reports the OLS estimates of β_1 and β_3 from the following global panel regressions:

$$RET_{i,t:t+n} = \alpha + \beta_1 SII_{i,t-1} + \beta_2 Recession_{i,t-1} + \beta_3 SII \times Recession_{i,t-1} + \beta_4 RET_{i,t-n:t-1} + \varepsilon_{i,t}\,,$$

where $RET_{i,t:t+n}$, $RET_{i,t-n:t-1}$, and $SII_{i,t-1}$ are defined as in Table 3. $Recession_{i,t-1}$ is a dummy variable that equals one if country *i* experiences a technical recession in week *t–1*, and zero otherwise. A country is said to experience a technical recession if its GDP decreases for two consecutive quarters (both quarters are assigned with a *Recession* dummy equal to one). In columns (4)–(6), I run similar regressions but replace *SII* with its non-detrended version, *EWSI*, defined as in Table 1. *SII* and *EWSI* are standardized to have a mean of zero and a standard deviation of one within each country. All coefficient estimates are in basis points. The sample period is from July 2006 to December 2016. Figures in brackets are *t*-statistics based on Driscoll-Kraay standard errors that are robust to cross-country correlations, heteroskedasticity, and autocorrelation. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	Independent variable: SII_{t-1}			Independent variable: $EWSI_{t-1}$		
	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$	$RET_{i,t:t+4}$	$RET_{i,t:t+8}$	$RET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)
SII or *EWSI*	-38.50^{***}	-95.78^{***}	-147.91^{***}	-33.93^{***}	-89.75^{***}	-139.32^{***}
	[–2.63]	[–4.00]	[–5.15]	[–2.60]	[–4.22]	[–5.54]
SII or *EWSI* × *Recession*	-150.90^{***}	-312.41^{***}	-397.00^{***}	-153.36^{***}	-302.00^{***}	-388.07^{***}
	[–3.29]	[–3.90]	[–4.18]	[–3.88]	[–4.24]	[–4.70]
NObs	16,752	16,624	16,498	16,843	16,715	16,589
Adj. R^2	2.05%	4.39%	6.53%	2.04%	4.27%	6.49%

Table A7

Statistics on stocks with different levels of short-selling activity

The table provides statistics on stocks that are split into terciles based on their weekly level of short interest (*SI*). The first three columns show the proportion of stocks that belong to the same tercile for 4, 8, and 12 consecutive weeks. The next three columns show percentage proportions that the stocks in each tercile, in aggregate, contribute to the market portfolio. The calculations for these proportions, *%cumMCap*, are identical to the ones described in Fig. 1. The sample period is from July 1, 2006 to December 31, 2016.

Country	% stocks in the same *SI* tercile for:			*Terc1* mean *%cumMCap*	*Terc2* mean *%cumMCap*	*Terc3* mean *%cumMCap*
	4 weeks	8 weeks	12 weeks			
Australia	88.55%	81.16%	75.09%	2.87%	34.21%	62.93%
Austria	85.94%	78.14%	71.69%	9.66%	27.13%	63.20%
Belgium	84.61%	75.81%	68.86%	6.33%	21.40%	72.27%
Brazil	77.32%	64.83%	55.82%	21.91%	46.74%	31.36%
Canada	86.86%	79.67%	73.84%	2.81%	23.39%	73.80%
Denmark	87.71%	80.47%	74.80%	4.36%	18.77%	76.87%
Finland	87.35%	80.22%	74.45%	2.28%	16.49%	81.24%
France	89.21%	81.84%	75.90%	3.97%	9.81%	86.22%
Germany	90.28%	83.70%	78.22%	3.33%	13.36%	83.31%
Greece	79.09%	68.11%	59.17%	23.40%	21.31%	55.29%
Hong Kong	90.49%	83.83%	78.35%	4.11%	30.69%	65.20%
Ireland	79.06%	67.38%	58.65%	5.31%	27.61%	67.08%
Israel	79.53%	68.93%	60.75%	14.25%	44.98%	40.78%
Italy	84.69%	75.34%	67.84%	3.54%	16.39%	80.08%
Japan	86.94%	77.64%	70.30%	5.18%	39.94%	54.89%
Korea	88.02%	79.19%	72.11%	27.43%	32.90%	39.67%
Mexico	92.61%	86.86%	82.30%	14.37%	55.49%	30.14%
Netherlands	85.83%	78.08%	71.92%	19.57%	36.18%	44.24%
New Zealand	82.89%	73.75%	66.62%	10.64%	31.02%	58.34%
Norway	83.31%	75.30%	69.03%	3.94%	15.45%	80.61%
Poland	86.95%	78.96%	72.55%	10.74%	32.82%	56.44%
Portugal	84.84%	76.30%	69.82%	4.15%	22.98%	72.87%
Russia	83.88%	75.26%	68.61%	28.77%	41.56%	29.67%
Singapore	82.46%	71.66%	63.13%	5.82%	16.80%	77.37%
South Africa	90.22%	83.24%	77.65%	4.81%	34.26%	60.93%
Spain	85.66%	76.71%	69.60%	5.47%	26.30%	68.23%
Sweden	86.58%	78.95%	72.88%	1.61%	12.81%	85.59%
Switzerland	86.77%	78.70%	72.49%	3.84%	15.92%	80.24%
Taiwan	84.92%	77.03%	70.53%	22.82%	36.02%	41.16%
Turkey	87.59%	78.70%	71.43%	11.02%	31.78%	57.21%
United Kingdom	83.88%	74.20%	66.32%	1.48%	47.74%	50.78%
United States	87.48%	79.63%	73.26%	47.77%	34.48%	17.75%
Average	85.67%	77.17%	70.44%	10.73%	29.05%	60.22%

Table A8

Cross-predictability effects in stocks with different levels of short interest

The table reports the OLS estimates of β_1 and β_3 from the following global panel regressions:

$$TercX\ \%NegativeEarnNews_{i,t} = \alpha + \beta_1 Terc1\ EWSI_{i,t-1} + \beta_2 Terc2\ EWSI_{i,t-1} + \beta_3 Terc3\ EWSI_{i,t-1} + \beta_4 Terc1\ RET_{i,t-4:t-1} + \beta_5 Terc2\ RET_{i,t-4:t-1} + \beta_6 Terc3\ RET_{i,t-4:t-1} + \varepsilon_{i,t}\,,$$

where *TercX %NegativeEarnNews*$_{i,t}$ denotes either the proportion of stocks in the portfolio tercile X that experience downward revisions in analysts' EPS forecasts from months *t–1* to *t*, *%Downgrades*, or the proportion of stocks in the portfolio tercile X that experience negative earnings surprises in month *t*, *%NegativeEAs* (all data are from I/B/E/S). EPS forecast revision in month *t* is defined as in Akbas et al. (2017) and equals to the mean EPS forecast in month *t* minus the mean EPS forecast in month *t–1* standardized by the absolute value of the mean EPS forecast in month *t–1*. Downward EPS forecast revision is a negative EPS forecast revision. Earnings surprise is defined as in Hirshleifer et al. (2009) and equals to the actual quarterly EPS released in month *t* minus the median of the most recent EPS forecasts for that quarter standardized by the end-of-quarter stock price. All estimates are based on quarterly EPS figures. *TercX EWSI* and *TercX RET* are defined as in Table 5. *TercX EWSI*s are standardized to have a mean of zero and a standard deviation of one within each country. Greece, Russia, and Turkey are omitted from the sample due to the lack of variation in *Terc1 EWSI*.or lack of earnings data in I/B/E/S. The sample period is from July 2006 to December 2016. Figures in brackets are *t*-statistics based on Driscoll-Kraay standard errors that are robust to cross-country correlations, heteroskedasticity, and autocorrelation. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively

	Dependent variable: *Terc1 %NegativeEarnNews*$_{i,t}$		Dependent variable: *Terc2 %NegativeEarnNews*$_{i,t}$		Dependent variable: *Terc3 %NegativeEarnNews*$_{i,t}$	
	Downgrades	*NegativeEAs*	*Downgrades*	*NegativeEAs*	*Downgrades*	*NegativeEAs*
	(1)	(2)	(3)	(4)	(5)	(6)
Terc1 EWSI	0.01	–0.02*	0.00	–0.01*	0.01*	–0.00
	(1.25)	(–1.97)	(1.05)	(–1.82)	(1.78)	(–0.34)
Terc2 EWSI	0.00	0.02	–0.02***	–0.00	–0.02***	–0.00
	(0.19)	(1.25)	(–3.21)	(–0.41)	(–4.35)	(–0.16)
Terc3 EWSI	0.00	0.02	0.01*	0.03***	0.00	0.01
	(0.50)	(1.19)	(1.94)	(2.90)	(0.86)	(0.85)
NObs	1,238	483	2,533	791	2,958	1,247
Adj. R^2	0.84%	4.54%	1.57%	2.49%	1.74%	1.82%

Table A9

Returns from the trading strategy

Panel A reports the OLS estimates of β_1 from the following global panel regressions:

$$FutRET_{i,t:t+n} = \alpha + \beta_1 TradingSignal_{i,t-1} + \varepsilon_{i,t},$$

where $FutRET_{t:t+n}$ is the cumulative return of the primary market index futures in a given country for n weeks starting from week t. $TradingSignal_{i,t-1}$ equals one if the equal-weighted average short interest (*EWSI*) across stocks in week $t-1$ is at least one standard deviation higher than its historical value, and zero otherwise. I require at least 100 weekly historical observations to obtain *TradingSignal*. In Panel B, I run similar regressions but replace $FutRET_{i,t:t+n}$ with strategy excess returns, $ExRET_{i,t:t+n}$, defined as the return on the investment strategy that sells the local market index futures and buys the local three-month government bonds. All coefficient estimates are in basis points. The sample period is from July 2006 to December 2016. In Panel A, the sample consists of 27 countries that have valid data on futures in Datastream. In Panel B, the sample consists of 20 countries that have valid data on futures and bonds in Datastream. Figures in brackets are t-statistics based on Driscoll-Kraay standard errors that are robust to cross-country correlations, heteroskedasticity, and autocorrelation. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Futures returns following the trading signal

	$FutRET_{i,t:t+1}$	$FutRET_{i,t:t+2}$	$FutRET_{i,t:t+3}$	$FutRET_{i,t:t+4}$	$FutRET_{i,t:t+8}$	$FutRET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)
TradingSignal	–21.76**	–46.18***	–73.58***	–97.62***	–183.89***	–256.35***
	[–2.40]	[–2.71]	[–3.04]	[–3.12]	[–3.30]	[–3.30]
Intercept	15.30	29.65	45.12	59.65	113.43*	163.54*
	[1.43]	[1.43]	[1.50]	[1.54]	[1.72]	[1.94]
NObs	14,385	14,359	14,333	14,307	14,203	14,099
Adj. R^2	0.07%	0.16%	0.26%	0.35%	0.61%	0.79%

Panel B: Investment strategy excess returns following the trading signal

	$ExRET_{i,t:t+1}$	$ExRET_{i,t:t+2}$	$ExRET_{i,t:t+3}$	$ExRET_{i,t:t+4}$	$ExRET_{i,t:t+8}$	$ExRET_{i,t:t+12}$
	(1)	(2)	(3)	(4)	(5)	(6)
TradingSignal	20.82**	42.66**	71.14***	98.69***	186.64***	265.11***
	[2.24]	[2.46]	[2.87]	[3.09]	[3.42]	[3.59]
Intercept	–12.58	–24.35	–38.56	–51.81	–96.39	–138.51
	[–1.15]	[–1.15]	[–1.26]	[–1.31]	[–1.43]	[–1.60]
NObs	10,525	10,505	10,485	10,465	10,385	10,305
Adj. R^2	0.07%	0.15%	0.26%	0.38%	0.67%	0.89%

Appendix B

To make sure that the Stambaugh (1999) bias does not drive my results in Table 3, I conduct a simulation similar to Kothari and Shanken (1997), Ang and Bekaert (2007), and Yu (2011). I discuss the results of these simulations in Section 4.2 and report them in Table A4 of Appendix A.

I focus on a univariate rather than a multivariate model to make fewer assumptions in simulations. Consider the following VAR model for market returns and its determinant, *SII*:

$$RET_{t:t+n} = \alpha_1 + \beta_1 SII_{t-1} + u_t,\ u \sim \text{i.i.d}\ (0, \sigma_u^2); \tag{B1}$$

$$SII_t = \alpha_2 + \beta_2 SII_{t-1} + \varepsilon_t,\ \varepsilon \sim \text{i.i.d}\ (0, \sigma_\varepsilon^2). \tag{B2}$$

RET and *SII* are defined as in Eq. (4), Section 4.1. I provide the OLS estimates of β_1 for all countries in my sample in columns (1)–(3) of Panel A, Table A4. The Stambaugh (1999) bias arises when u and ε are correlated. This correlation biases the OLS estimate of β_1 in Eq. (B1) and inflates its statistical significance.

I conduct two types of simulations to investigate how the Stambaugh (1999) bias affects my results. First, I gauge the magnitude of a potential Stambaugh (1999) bias in estimated β_1. Following Kothari and Shanken (1997), I start my simulations with the first historical value of *SII* in my sample period. Given the actual estimates of β_1 and β_2 in Eq. (B1)–(B2), I simulate the values of *RET* and *SII* by randomly drawing a pair of residuals (u, ε) with replacement. I then use the simulated time-series of *RET* and *SII* to estimate β_1 in Eq. (B1). For each country, the procedure is repeated 10,000 times. The difference between the average simulated β_1 and the actual β_1 (reported in columns (1)–(3) of Panel A, Table A4) constitutes the estimated Stambaugh (1999) bias, which is shown in columns (4)–(6) of Panel A, Table A4.

Second, I assess whether the correlation between u and ε can undermine my statistical inferences. Specifically, I test how likely I can obtain the similar β_1 coefficients and *t*-statistics as

I observe in the data if the "true" coefficients are actually zero. Again, I start my simulations with the first historical value of *SII* in my sample period. Given that $\beta_1 = 0$ and β_2 is equal to its actual estimate in Eq. (B2), I simulate the values of *RET* and *SII* by randomly drawing a pair of residuals ($\tilde{u}$, ε) with replacement, where $\tilde{u}$ denotes the residual from the model: $RET_{t:t+n} = \tilde{\alpha}_1 + \tilde{u}_t$, $\tilde{u}$ ~ i.i.d $\left(0, \sigma_{\tilde{u}}^2\right)$. I then estimate β_1 in Eq. (B1) and its corresponding Newey-West *t*-statistics 10,000 times. Finally, I compute the *p*-values for the actual slope coefficients in columns (1)–(3) of Panel A by comparing them to the percentiles of the simulated distribution of β_1 and its *t*-statistics. In particular, I look at the proportions of the simulated β_1 and its *t*-statistics that are lower (more negative) than the actual β_1 and its *t*-statistics, respectively. These *p*-values represent the probability under the null ($\beta_1 = 0$) of obtaining a historical slope coefficient and its *t*-statistics as negative as the ones observed in the actual data. The upper part of Panel A in Table A4 summarizes the results. Statistical inferences in columns (1)–(3) are based on *p*-values drawn from the comparison of the actual and simulated *t*-statistics.

Similar procedures are run for the pooled global panel regressions in Panel B of Table A4. I preserve country-specific coefficients in Eq. (B2) but set all slope coefficients in Eq. (B1) to be equal across countries.