HONG KONG INSTITUTE FOR MONETARY AND FINANCIAL RESEARCH

TOWARD UNDERSTANDING SHORT-SELLING ACTIVITY: DEMAND AND SUPPLY

Adrian W.K. Cheung, Hung Wan Kot, Eric F.Y. Lam, and Harry K.M. Leung

HKIMR Working Paper No.17/2019

December 2019

Hong Kong Institute for Monetary and Financial Research
香港貨幣及金融研究中心
(a company incorporated with limited liability)

All rights reserved.
Reproduction for educational and non-commercial purposes is permitted provided that the source is acknowledged.

Toward understanding short-selling activity: Demand and supply *

Adrian W.K. Cheung
Flinders University

Hung Wan Kot
University of Macau

F.Y. Eric Lam
Hong Kong Institute for Monetary and Financial Research
Hong Kong Academy of Finance, Hong Kong Monetary Authority

Harry K.M. Leung
Cheung Kong Graduate School of Business

December 2019

Abstract

We investigate the demand and supply sides of short-selling activity in the U.S. from 2003 to 2015. We construct four types of demand side variables from fundamentals, and three types of supply side variables from institutional ownership (IO) and stock loan data. The supply side variables play a more important role in determining short selling than the demand side variables. The IO of quasi-indexer type is the most important supply side variable, while the arbitrage and hedging with options market is the most important demand side variable. Finally, a portfolio sorting approach confirms the same results.

Keywords: short selling, demand and supply, institutional ownership, borrowing cost
JEL code: G01, G11, G14

---

* Corresponding author: Cheung: Adrian.Cheung@flinders.edu.au, Kot: hungwankot@umac.mo, Lam: cfylam@hkma.gov.hk, Leung: lhkharry@gmail.com

For their helpful comments and suggestions, we would like to thank Kalok Chan, Sudeep Ghosh, Yeonjeong Ha, Zsuzsa Huszar, Joseph Lee (SFC Hong Kong), Yafeng Qin, Tom Smith (Editor), Lewis Tam, two anomalous referees, and the participants at the finance seminars at Curtin University, Fudan University, Hong Kong Polytechnic University, University of Macau, Massey University, University of South Australia, the 7th International IFABS Conference at Hangzhou in 2014, the Auckland Finance Meeting and the AsianFA 2015 Annual Meeting at Changsha. Kot acknowledges a start-up research grant from University of Macau (FRG2017-00103-FBA). This paper represents the views of the authors, which are not necessarily the views of the Hong Kong Monetary Authority, Hong Kong Institute for Monetary and Financial Research, its Council of Advisers, or the Board of Directors.

1. Introduction

This paper investigates the demand and supply side determinants of short-selling activity. Short selling plays an important role in financial markets, as it constitutes 24% of the NYSE trading volume and 31% of the Nasdaq trading volume (see, e.g., Diether, Lee and Werner, 2009a). Stock prices are more accurate when short sellers are more active in trading, and market quality declines when short selling is banned (see, e.g., Boehmer and Wu, 2013; Boehmer, Jones and Zhang, 2013; Helmes, Henker and Henker, 2017).

The literature has largely focused on a few prominent aspects of short selling. These aspects include the negative relationship between short selling and future stock returns,[1] whether short sellers possess informational advantages and are skilled investors producing abnormal returns,[2] the actual costs of short selling, the rebate rates of the borrowed stock,[3] the effects of institutional ownership (IO) on short-sale constraints,[4] institutional details[5], and the effects of short selling on real corporate activity.[6]

Our study extends the short selling literature by investigating the process by which short interest (i.e., the number of shares sold short) is determined. While previous studies examine links between future returns and the demand and supply factors of short selling, the link between these factors and short selling itself is less clear. Previous studies show that supply is not binding,

[1] For example, Chen, Hong and Stein (2002), Desai, Ramesh, Thiagarajan and Balachandran (2002), Nagel (2005), Boehme, Danielsen and Sorescu (2006) and Cohen, Diether and Malloy (2007).
[2] For example, Dechow, Hutton, Meulbroek and Sloan (2001), Christophe, Ferri and Angel (2004), Boehmer, Jones and Zhang (2008, 2012), Diether, Lee and Werner, (2009b), Engelbert, Reed and Ringgenberg (2012), Chen, Desai and Krishnamurthy (2013) and Jiao, Massa and Zhang (2016).
[3] For example, D'Avolio (2002), Geczy, Musto and Reed (2002), Ofek, Richardson and Whitelaw (2004), Kaplan, Moskowitz, and Sensoy (2013), and Kolasinski, Reed, and Ringgenberg (2013).
[4] For example, Nagel (2005), Asquith, Pathak and Ritter (2005) and Prado, Saffi and Sturgess (2016).
[5] For example, regulatory regime changes (Jain, Jain, and McInish, 2012), failure to deliver (Jain and Jain, 2015), and regulatory reach (Jian, Jian, McInish, and McKenzie, 2013). We thank an anomalous referee for the suggestion to include the aspects of institutional details.
[6] For example, Grullon, Michenaud and Weston (2015), Chang, Lin and Ma (2015), Jiang and Pang (2016) and Massa, Wu, Zhang and Zhang (2015).

given the low utilization rate.[7] If supply is in excess of demand, then demand is the binding factor shaping short selling. In addition, Hanson and Sunderam (2014) also find that among stocks with high institutional ownership, there is an upward trend in short interest, which suggests that shorting demand may have played an important role in driving the long-term trends of short interest.

We examine a large panel of firms on the NYSE/Amex/Nasdaq from 2003 to 2015 and construct four types of demand side variables from fundamentals, and supply side variables like ownership concentration, different types of financial institutions, and borrowing costs from the institutional ownership and stock loan data. We find that the supply side plays a more important role in determining short selling. The IO of quasi-indexer type is the most important supply side variable while the arbitrage and hedging with options market is the most important demand side variable. Our results are robust to alternative models and sub-samples. We also analyze the relations between changes in the determinants and the actual changes in short interest. We find that changes in supply side variables are more important than changes in demand side variables for determining the changes in short interest ratio (SIR, i.e., the percentage of outstanding shares sold short). Finally, to investigate whether the demand or the supply side is more important in determining short selling, we extract the common demand and supply factors from the principal component analysis, and sort portfolios based on these common factors. The results confirm that the supply side is more important than the demand side.

Previous studies use short interest as a proxy for shorting demand. However, as short interest is the equilibrium outcome of shorting demand and shorting supply (Asquith et al., 2005),

[7] Early studies such as D'Avolio (2002) report that the aggregate market is easy to borrow and that most stocks can be borrowed. Asquith et al. (2005) classify approximately 21 stocks per month as short-sale constrained, and suggest that short-sale constraints are not common. Recently, Prado et al. (2016) report that lending supply was around 20-25% for most of the period, while the utilization rate of shares on loan was around 14.6-24% of lending supply.

it is difficult to isolate the impact of shorting supply from that of shorting demand if short interest is used as a proxy for shorting demand. We overcome this difficulty by explicitly modelling shorting demand from demand-side factors such as: derivative arbitrage/hedging; difference between investor opinions; overvaluation and momentum trading. Our study complements that of Cohen et al. (2007), who rely on the fees and quantity of four year stock loans to isolate the shifts in shorting demand from that of shorting supply, and to test how shorting impacts future returns. We focus on the shorting demand from the original fundamental sources, to proxy for the demand of short selling. These sources are overvaluation, momentum, arbitrage and hedging demand from the options market, and difference in investor opinions. Our study also complements that of Kot (2007), which tested mainly hypotheses of short selling related to the demand side factors. We use supply information from both IO and stock loan data to provide a more complete picture of the supply side.

In line with Nagel (2005) and Asquith, Pathak and Ritter (2005), we use IO as the standard measure of shorting supply, and find similar results. Following Bushee (1998, 2001), we further break down ownership into dedicated, quasi-indexer, and transient institutional investors and find that the supply of shares from the quasi-indexer has the largest effect among supply side variables. The results are consistent with Prado et al. (2016), who show that IO by long-term investors further raises the shorting supply. We also echo Prado, Saffi and Sturgess (2016), who find that ownership concentration reduces shorting supply and, hence, short interest. We show that high concentration of ownership reduces short interest.

Our paper also complements recent studies on the supply side of short selling. Kaplan et al. (2013) conduct an experiment where they take high loan fee stocks in a manager's portfolio and randomly make available, and withhold, these stocks from the lending market, creating an exogenous and sizeable shock to the supply of lendable shares. Supply shocks like this are expected to reduce market lending fees and raise quantities, significantly. However, Kaplan et al. (2013) find no evidence that stock returns, volatility, skewness, or bid-ask spreads are affected. Kolasinski et al. (2013) use unique data from 12 lenders to examine how equity lending fees respond to demand shock. They find that increases in shorting demand increase lending fees significantly only when the current level of demand is already high. Our study examines a larger panel dataset with a longer sample period, using both regression and portfolio sorting approaches, to address the effects of shorting demand and supply. Our key finding is that supply-side variables consistently play a more important role than demand-side variables in determining the patterns of short selling.

Finally, Boehme, Danielsen and Sorescu (2006) suggest that idiosyncratic stock return volatility is a proxy for divergence in investor opinions, which, in turn, raises shorting demand in anticipation of overvaluation.[8] However, Prado, Saffi and Sturgess (2016) suggest that idiosyncratic volatility, as a proxy for arbitrage risk, increases the difficulty of short selling. We verify, empirically, that idiosyncratic volatility raises short interest, suggesting that the overvaluation effect dominates the arbitrage risk effect.

The remainder of this paper is organized as follows. Section 2 details the data sources and the construction of variables. Section 3 provides the results of the regression analysis, including sub-period analysis and changes in SIR. Section 4 provides the results of the portfolio sorting analysis. Section 5 presents our conclusions.

[8] The intuition is that, given short-sale constraints, investors are not able to fully incorporate negative information into stock prices as proposed by Miller (1977). For positive information, however, there is no such restriction.

2. Methods

2.1 Data

The monthly short-interest data are obtained from Compustat for July 2003 to December 2015.[9] Our sample starts from July 2003 because Compustat covers all NYSE/Amex/Nasdaq short interest from July 2003. We focus on the common stocks (share code 10 or 11) in NYSE/Amex/Nasdaq only, i.e., ADR, REIT, and ETF are not included. The stock loan data are obtained from Markit equity lending database.[10] The daily and monthly stock return data, daily closing bid-ask spreads, and CRSP value-weighted index are retrieved from CRSP. The accounting data are obtained from the annual Compustat. The quarterly IO data are obtained from Thomson Financial's institutional database. The classification of institution types is obtained from the website of Bushee.[11] The information on options availability is obtained from the Option Metrics. The analysts' earnings forecast data are obtained from the I/B/E/S summary file. The S&P 500 Index constituents are taken from Compustat's monthly updates.

We merge short-interest data with CRSP/Compustat by Cusip. Over our sample period, on average, 90.2 percent of NYSE/Amex/Nasdaq common stocks have reported short interest. We treat other stocks without reported short interest as having zero short interest.

Figure 1 shows the time-series for the number of stocks that have reported short interest, and the monthly average of SIR, on the NYSE/Amex and Nasdaq from July 2003 to December

---

[9] Short interest has been published in the middle, and at the end, of each month since 2007. It had previously been published only in the middle of the month. For consistency, the short interest recorded in the middle of each month was used for the entire period.
[10] Markit stock loan data starts from June 2006, but with low NYSE/Amex/Nasdaq stocks coverage. For example, the coverage for stock loan utilization variable is 48 percent in June 2002, but increases to 73 percent in July 2003.
[11] http://acct.wharton.upenn.edu/faculty/bushee/IIclass.html. Bushee mentions that the 2016 Spectrum data has significant errors and he will not be updating for 2016 until these problems have been resolved. The related data issue is also discussed in WRDS.

2015. The SIR increases significantly before the financial crisis in 2008. The SIR is from 2.78 percent in July 2003 to 7.25 percent in July 2008, then drops to 4 percent in January 2009. In the post-crisis period, the SIR appears to be quite stable. Figure 1 begs an interesting question: do the demand and supply sides play different roles in determining the SIR in the pre-2010 volatile and post-2010 stable periods?

[Figure 1 here]

Stock loan data is also merged with CRSP/Compustat by Cusip. Three variables are used and they are stock loan utilization (UTL), daily cost of borrow score (DCBS), and stock loan fee (FEE). Among the three variables, UTL has the highest coverage: on average 89.6 percent of all NYSE/Amex/Nasdaq stocks have reported UTLs. The corresponding coverage of DCBS is 82.9 percent over the sample period. FEE has the lowest coverage, with the corresponding coverage at only 48.6 %. The FEE data starts from April 2007 only: on average the coverage is only 12 percent in the 2003-2009 period, while the coverage is 88 percent in 2010-2015 period.

Bushee provides annual updates of institution classifications on his website. We merge Bushee's data with quarterly Thomson Financial's institutional database, to compute the institutional holding of each type of institution at stock level. Then we merge the holding data with the monthly number of shares outstanding from CRSP, to construct the monthly IO holding of various institutions. Following Prado, Saffi, and Sturgess (2016), we also construct the ownership concentration variables (IO_HHI and IO_TOP5) on a monthly basis.

2.2.1 Demand side factors

Recall that four types of demand for short selling at stock level are examined, and they are stock overvaluation, arbitrage and hedging, momentum trading, and difference in investor opinions. To capture stock overvaluation, we follow Stambaugh, Yu, and Yuan (2015), to construct a variable called the relative valuation variable (RV). This RV varies from 1 to 10. Note that higher RVs predict higher returns. Therefore, a negative relationship between RV and SIR is expected.[12]

To capture the demand from arbitrage and hedging activities, we use D_OPT, a monthly dummy variable that takes a value of 1 if some options are written on the firm's stock, and otherwise zero. Market makers in options markets must short stocks to hedge their positions, and some investors may take a short position of stocks if the options price deviates from put-call parity (Grundy et al., 2012; Hu, 2014). Hence, D_OPT is expected to be positively related to SIR.

To capture the demand from momentum trading activities, we use PR1Y. This is the cumulative return on a firm's stock in the previous year from month *t*-12 to *t*-1. Jegadeesh and Titman (1993) document a price continuation effect over a 3- to 12-month period, and show that a momentum strategy of buying past winners and selling past losers can earn a return of 1% per month. Some short sellers are trend traders, who close their position if the stock price has increased recently and open a new short position if it has decreased. This momentum strategy suggests a negative relationship between PR1Y and SIR.

Finally, idiosyncratic risk (IDIO) is used as a proxy for the range of investor opinions (Boehme et al., 2006). It is measured as the mean squared error of residuals of daily stock returns

[12] We acknowledge a referee for the suggestion to use RV. The RV variable is constructed by combining 11 return anomalies, including financial distress, net stock issues and composite equity issues, total accruals, etc. For details of how the RV variable is constructed and summary statistics of this variable, please refer to an internet appendix available on the website of *Accounting and Finance*. In addition, in previous versions, we simply use book-to-market ratio (BM), and industry adjusted-BM, to proxy for overvaluation (Dechow et al., 2001; Nagel, 2005). Though not reported, we find a negative relationship of BM with SIR (same for industry adjusted-BM).

within last three months, from Carhart (1997)'s 4-factor model. Given the same level of short-sale constraints, more investors will short a stock if there is a large range of opinions. Boehme et al. (2006) show that a dispersion in investor opinions in the presence of short-sale constraints leads to stock price overvaluation. This suggests a positive relationship between IDIO and SIR.[13]

2.2.2 Supply side factors

As the majority of lendable shares are from institutional investors, we use institutional ownership (IO), which is defined as the fraction of a firm's outstanding shares owned by institutions on a monthly basis, to proxy for the short-selling supply. Apart from IO, we also constructed two sets of IO variables.

Prado, Saffi, and Sturgess (2016) show that ownership concentration is related to lending supply — that is, higher concentration stocks exhibit lower lending supply. The reason is that investors with larger stakes may prefer to withhold their shares from lending in order to maintain greater ability to monitor management, because the value of their holdings could be lowered by short selling (Prado, Saffi, and Sturgess, 2016). Following Prado, Saffi, and Sturgess (2016), we construct two proxies, namely, IO_TOP5 and IO_HHI. The former is defined as the ownership of the largest five institutional shareholders scaled by the total institutional ownership. The latter is the IO concentration by the Hirschman-Herfindahl index. Given that IO_TOP5 and IO_HHI are associated negatively with the equity lending supply, we expect that IO_TOP5 and IO_HHI are also related negatively to SIR.

[13] Another popular proxy is the dispersion in analysts' earnings forecasts (DISP), which requires that the stock be covered by at least three analysts. In a robustness check reported in the paper, we also include DISP as one of the demand side variables.

Following Bushee (1998, 2001), we construct IOs for three different types of institutions: IH_TRA, IH_DED, and IH_QIX. IH_TRA are the transient institutions that have high portfolio turnover and diversified portfolios. IH_DED are the dedicated institutions that have low turnover and more concentrated portfolio holdings. IH_QIX are the quasi-indexer institutions that have low turnover and diversified portfolio holdings (consistent with many index strategies). This approach can capture the popularity and significant growth of the index fund industry (including index funds and ETFs) during the sample period, and provides the largest supply in stock loan markets. We expect all of IH_TRA, IH_DED, and IH_QIX to be related positively to SIR, and the impact of IH_QIX on SIR to be greater than those of IH_TRA and IH_DED.

Finally, we construct three variables from the stock loan data: DCBS, UTL, and FEE. DCBS is the monthly average of the daily cost of borrow score, 1 for the cheapest and 10 for the most expensive to borrow. UTL is the stock loan utilization, defined as beneficial owners (BO) on loan value divided by BO inventory value. FEE is the monthly average of the daily borrowing fee. Those variables have been used extensively in previous studies (e.g., Beneish et al., 2015; Prado et al., 2016).[14] In general, the lending fee still has a positive relationship with SIR. Therefore, we expect DCBS, UTL and FEE to be related to SIR positively.

2.3 Control variables

Short sellers take liquidity risk into account, and greater liquidity in a stock should enhance short selling activity. We measure liquidity by the bid-ask spread (SPREAD) and Amihud's (2002) illiquidity measure (ILLIQ). SPREAD is a monthly average of the daily

[14] Although short-sale constraints are not binding on most stocks, i.e., the supply is more than the demand, D'Avolio (2002) finds that the shorting supply curve is kinked, i.e., the lending fee is inelastic for low SIR levels and has a positive relationship with SIR beyond the kink.

closing bid-ask spread, which is calculated as the bid-ask difference scaled by the average of the bid and ask stock prices. ILLIQ is the measure of stock illiquidity developed by Amihud (2002) from the daily stock price and trading volume. These two variables measure different aspects of liquidity. SPREAD captures the cost of immediacy, and ILLIQ captures how much a given trading volume moves the price (Corwin and Schultz, 2012). In either case, we expect the relationships between illiquidity and SIR to be negative.

Finally, we include LNCAP, which is the natural logarithm of a firm's market capitalization calculated monthly, and D_SP500, which is a dummy variable indicating whether the stock is included in the S&P 500 Index in a given month (1 for yes, 0 for no). These are standard controls.

2.4 Preliminary statistics

Table 1 reports the summary statistics. All non-dummy variables are winsorized at the bottom and top 1 percent levels. The average SIR is 3.78 percent among all stocks. The pattern and magnitude are close to those reported in Hanson et al. (2014). [15] The average of D_DEBT and D_OPT are 0.16 and 0.53, respectively. This means that, on average, 16 percent of the sampled stocks have issued convertible debts and 53 percent of them have listed stock options. The average PR1Y is 15.31 percent, meaning that there is an upward trend over the sample period.

Among the supply side variables, the average of IH_DED, IH_QIX, and IH_TRA are 3.1 percent, 32.8 percent, and 11.7 percent, respectively. This means that most institutions are quasi-

[15] The average of SIR reported in the Beneish et al. (2015) paper is 5.2 percent over the 2004-2013 period. The difference is mainly due to the fact that Beneish et al. exclude firms with missing short interest and stock loan data, as well as financial institutions and utilities firms.

indexer institutions, and many of their trading strategies are consistent with index strategies. Though not reported, we further check the time-series changes of IH_DED, IH_QIX, and IH_TRA during the sample period. IH_DED and IH_TRA are relatively stable: the averages of IH_DED are 6 and 7.5 percent at the beginning and ending of the sample period while the corresponding averages of IH_TRA are 11.9 and 13 percent. However, we find a marked upward trend in IH_QIX during the sample period, and the averages of IH_QIX are 18.8 and 31.9 percent at the beginning and end. The Bushee's classification of institution types captures the rapid increase of the index fund markets, including both index funds and ETFs. For the ownership concentration measure, the top five holdings (IO_TOP5) count for 58 percent of all institutional holdings. This number is close to the IO_TOP5 reported in Prado et al. (2016): 54.56 percent.

As far as the three stock loan variables are concerned, the average DCBS, UTL, and FEE are 1.55, 18 percent, and 1.85 percent respectively. The magnitude of the average UTL is close to that reported in Prado et al. (2016), which is 19.25 percent. Though Prado et al (2016) do not use DCBS in their study, we find the magnitude of the average DCBS in our sample is close to the number reported in Beneish et al. (2015), which is 1.642. The average level of FEE in our sample is significantly higher than that reported in Prado et al. (2016), which is 0.71 percent.[16]

[Table 1 here]

[16] The difference is mainly due to the fact that different variables are used. The Markit database provides an SAF in the 2003-2009 period and an indicative fee in the 2010-2015 period. Prado et al. (2016) use the borrow cost from hedge funds in the stocks (variable SAF in the database), which is available only in the 2003-2009 period. We combine them together to construct the variable. Note that the average SAF in our 2003-2009 period is 0.71 percent, which is same as that of Prado et al. (2016).

Table 2 reports the Pearson's correlations among variables.[17] The signs of all demand, supply and control variables correlated with SIR are consistent with predictions. Market capitalization (LNCAP) has high correlations with stock options availability (D_OPT), idiosyncratic volatility (IDIO), institutional ownership (IO, IO_HHI, IO_TO5), and the dummy of S&P 500 index constituent stocks (D_SP500). D_OPT is also highly correlated with IO and IO_TOP5. Among the institutional ownership variables, both IH_QIX and IH_TRA are highly correlated with IO and IO_TOP5. In addition, IO, IO_HHI, and IO_TOP5 are also highly correlated. We shall pay attention to these variables to run the multivariate regressions, as they may create a multicollinearity problem.

[Table 2 here]

3. Regression analysis

3.1 Multivariate regressions

In this sub-section, we apply the multivariate regression analysis to study the impact of demand and supply factors on the SIR by using the following pooled OLS regression with monthly data.

$$
\begin{aligned}
SIR_{i,t} = {} & \alpha + \beta_1 RV_{i,t} + \beta_2 D\_OPT_{i,t} + \beta_3 PR1Y_{i,t} + \beta_4 IDIO_{i,t} \\
& + \gamma_1 IO\_HHI_{i,t} + \gamma_2 IH\_DED_{i,t} + \gamma_3 IH\_QIX_{i,t} + \gamma_4 IH\_TRA_{i,t} + \gamma_5 DCBS_{i,t} + \gamma_6 UTL_{i,t} \\
& + \varphi_1 D\_SP500_{i,t} + \varphi_2 SPREAD_{i,t} + d_{m,t} + \varepsilon_{i,t} \qquad (1)
\end{aligned}
$$

[17] We drop DISP and FEE in order to have more observations to compute the correlations. If we include DISP and FEE, the observations will drop from 413190 to 153663. If we include DISP and FEE, we find FEE has high correlations with DCBS (0.82).

To absorb the time-invariant stock-specific effects and time-varying effects, we include stock fixed-effects(f_i) and year-month dummy variables $\left(d_{m,t}\right)$ in the regression model.

Based on equation (1), four models are estimated and their results are reported in Table 3. All models include stock fixed-effects and time fixed-effects. Standard errors are double clustered at the firm and year-month levels. In each month, we standardize all of the variables to have zero mean and unit standard deviation (Petersen, 2009; Prado, Saffi & Sturgess, 2016). The advantage is that we can compare the magnitude of coefficients directly in models 1-4. We exclude LNCAP because it is highly correlated with many other variables. We also exclude IO_TOP5 and ILLIQ, as the former is highly correlated with IO_HHI, while the latter is highly correlated with SPREAD. We do not include DISP and FEE in the regression as these two variables have much fewer observations. However, we will provide robustness checks of those variables in next sub-section.

Model 1 of Table 3 reports the results for the five demand-side variables. All of them have significant coefficients, except PR1Y. Consistent with the literature, the coefficient of D_OPT is positive and statistically significant, suggesting that arbitrage and hedging demand drives short selling from the stock options. The results are consistent with previous studies. For example, Pan and Poteshman (2006) find that the option trading volume contains information about future stock prices. Blau and Wade (2013) compare short sales and put options in terms of return predictability, and find that the predictability of short sales is higher than that of put options. Blau and Wade also find that investors who face short sale constraints will shift to take short positions (i.e., buying put options) in the options market instead. Hu (2014) also documents that after executing option orders, options market makers turn to the stock market to hedge away the underlying stock exposure.

The negative coefficient of RV means that short sellers prefer to short stocks that have lower relative valuation, and avoid short selling stocks that have higher expected returns. The positive coefficient of IDIO means that, given the same level of short-sale constraints, short sellers short more if the underlying stocks show a large range of investor opinion. However, Prado, Saffi and Sturgess (2016) argue that idiosyncratic volatility, as a proxy for arbitrage risk, increases the difficulty of short selling. We verify empirically that idiosyncratic volatility raises short interest, suggesting that the overvaluation effect dominates the arbitrage risk effect.

[Table 3 here]

Model 2 of Table 3 reports the results of the supply-side variables. All variables are statistically significant. The negative coefficient of IO_HHI is consistent with Prado et al. (2016). The results show that ownership concentration is related to lending supply: high concentration stocks exhibit low lending supply. The reason is that investors with large stakes may prefer to withhold their shares from lending, to maintain greater ability to monitor management, because the value of their holdings could be lowered by short selling (Prado, Saffi, and Sturgess, 2016). Consistent with the idea that high concentration of ownership comes with low lending supply, Table 2 shows that IO_HHI is correlated positively with DCBS and correlated negatively with UTL.

Among the three types of IO classified by Bushee's method, all of them are related to SIR positively. Consistent with our conjecture, the magnitude of IH_QIX is higher than those of IH_DED and IH_TRA. The corresponding coefficients of IH_QIX, IH_DED, and IH_TRA are 0.34, 0.04, and 0.19, respectively. The results suggest that quasi-indexer type institutional investors provide the largest portion of stock supply in the loan markets, while the transient type institutions also provide a significant portion of stock supply.[18] Our results are also consistent with D'Avolio (2002) and Prado, Saffi, and Sturgess (2016) in the sense that there is a significantly positive relationship between long-term holding institutions and SIR.

For the variables from the stock loan data, we also find positive relationships of DCBS and UTL with SIR. Consistent with Table 1, which shows that, on average, only 18 percent of stocks on loan are borrowed, the results show that most stocks are not bound by short-sale constraints. Higher DCBS and UTL means that whenever more short-sellers want to short stocks, that higher demand is also reflected into the supply side of stock loan markets. Positive coefficients for DCBS and UTL are consistent with Diamond and Verrecchia (1987), who argue that imposing a cost on short-selling obviously makes it less attractive, and one expects that those willing to pay the cost are the ones with the greatest anticipated benefits from selling short.

Model 3 of Table 3 estimates demand and supply variables together, and Model 4 further adds D_SP500 and SPREAD as control variables. As the results of demand and supply variables are similar in these two models, we focus our discussion on Model 4. Among the 13 explanatory variables, the largest five coefficients are from IH_QIX, UTL, DCBS, IH_TRA, and D_OPT, and the corresponding values are 0.313, 0.258, 0.248, 0.177, and 0.078, respectively. All these are supply side variables except D_OPT. The total effect of these supply side variables is greater than that of the demand side variables. This means the supply side is more important than the demand side in determining the SIR. Taking IH_QIX as the example, a one-standard-deviation increase in quasi-indexer type institution ownership is associated with an SIR of 0.313 standard deviations higher, which is equivalent to a 44.6 percent (=0.313*0.0539/0.0378) increase relative

[18] For example, in note 7 to the financial statements of its 2012 annual report, the Fidelity Focused Stock Fund states that "the Fund lends portfolio securities through a lending agent from time to time in order to earn additional income … Total security lending income during the period amounted to $61,400." The portfolio turnover of the fund was 279% in 2012.

to the mean SIR. The corresponding value for DCBS is 35.3 percent (=0.248*0.0539/0.0378), for UTL 36.7 percent (=0.258*0.0539/0.0378), and for IH_TRA 25.2 percent (=0.177*0.0539/0.0378).

Among the demand side variables, the coefficients of D_OPT, IDIO, and RV are significant, and the corresponding values are 0.078, 0.047, and -0.031, respectively. A one-standard-deviation increase in D_OPT is associated with an SIR 0.078 of a standard deviation higher, which is equivalent to an 11.1 percent (=0.078*0.0539/0.0378) increase relative to the mean SIR. The corresponding value for IDIO is 6.7 percent (=0.047*0.0539/0.0378), and for RV -4.4 percent (=-0.031*0.0539/0.0378).

The negative coefficient of D_SP500 means that short-sellers avoid shorting S&P 500 index constituent stocks. The intuition is that those stocks are large firms, with much higher information transparency, more analyst coverage and more attention from investors. Therefore, the room for short-sellers to make profit from such stocks is much lower. The negative coefficient of SPREAD means that short sellers prefer to short stocks that have higher liquidity.

3.2 Alternative models and sub-sample analysis

In this sub-section, we provide two sets of robustness checks of Table 3. Panel A of Table 4 provides the estimation results of alternative models that include variables additional to those of the Table 3 models. We provide six models in Panel A. Model 1 of Panel A includes IO, which is the total ownership (IO) of stocks hold by institutional investors. The coefficient of IO is the highest among all variables (0.378 with a t-value of 17.11); higher than the coefficients of

DCBS (0.239 with a t-value of 13.9) and UTL (0.267 with a t-value of 13.63). Model one suggests that IO is good proxy for the supply side variables of short selling.

Model 2 of Panel A includes DISP as an additional demand side variable. The positive coefficient of DISP (0.013 with a t-value of 4.01) means that the dispersion of analysts' earnings forecasts is also a good proxy for the range of investor opinions. In Model 3, we replace IO_HHI by IO_TOP5; another proxy for the ownership concentration. Consistent with Prado et al. (2016), the coefficient is negatively significant (-0.123 with a t-value of 9.22). Model 4 replaces SPREAD by ILLIQ. However, the coefficient of ILLIQ is statistically insignificant after controlling for other variables, even though the univariate regression shows that ILLIQ is related statistically and negatively to SIR. Finally, we include FEE in models 5 and 6 in Panel A. The coefficient of FEE is positively significant, which is consistent with our conjecture.

[Table 4 here]

Panel B reports the multivariate regression results of sub-sample analyses of Model 4 of Table 3. First, we split the sample into the 2003-2008 and 2009-2015 periods. Recall from Figure 1, which shows that the SIR is volatile in the first sub-period, and remains relatively stable in the second sup-period. Given the fact that most stocks are free from short-sale constraints, as the supply of stocks in the loan market is greater than the demand (D'Avolio, 2002; Prado et a., 2016), we hypothesize that demand will play a more important role in the 1st sub-period, than in the 2nd sub-period. The first and second columns of Panel B confirm our conjecture. The coefficients of D_OPT and IDIO are higher in the 2003-2008 period. However, PR1Y is statistically significant and negative in the 2009-2015 period only. Since short sellers, and hedge funds in particular (see Goldman Sachs, 2009), change their trading strategies according to market conditions, they may adopt momentum or contrarian strategies. Though not tabulated, we re-estimate the Model 4 of Table 3 by year, and find that among the 12 years, the coefficient of PR1Y is statistically significant and negative in six years (2007, 2010, 2012-2015), significantly positive in one year (year 2009), and insignificant in the remaining five years. The results suggest that short sellers adopt momentum strategies more than contrarian strategies.

IO_HHI is statistically significant and negative only in the 2003-2008 period. We also further run the Model 4 of Table 3 by year to obtain the coefficient of IO_HHI. Among the seven years of 2009-2015, the coefficient of IO_HHI is significant and negative in 2009, 2014 and 2015. Though the overall coefficient of IO_HHI is insignificant in the 2^{nd} sub-period, ownership concentration still plays a role in affecting the short-selling activity negatively.

We also check for the distribution of SPREAD in the two sub-periods. The median value of SPREAD is 0.0033 and 0.0017 in the 1^{st} and 2^{nd} sub-periods, implying a 48 percent decrease from the 1^{st} to the 2^{nd} period. This is consistent with the view that the spread is of less concern to short sellers in the 2^{nd} sub-period. As such, the coefficient of SPREAD is insignificant.

Next, we investigate the difference between NYSE/Amex and Nasdaq. The results show that the coefficient of RV is -0.046 (with a t-value of 6.97), suggesting that short sellers only short overvalued stocks on the Nasdaq. In general, there are more technological stocks listed on the Nasdaq, while the valuation of those stocks is more difficult than that of traditional stocks listed on the NYSE/Amex. Short sellers, mainly hedge funds, have the advantage and expertise to value those stocks and take the short position. In addition, short-sellers play as momentum investors only on the NYSE/Amex, as the coefficient of PR1Y is -0.029 (with a t-value of 3.62).

In column five of Panel B Table 4, we report on whether the main result remains after excluding financial firms (SICCD codes between 6000 and 6999). We find similar results to those reported in the Model 4 of Table 3.

Finally, in column six of Panel B Table 4, we check for the "*special*" stocks, which are relatively difficult to borrow unless a high loan fee is paid. We are interested to know whether the demand and supply factors of these special stocks are different from those of non-special stocks. Following Beneish et al. (2015), we define a stock as special if DCBS is greater than or equal to 3. The results show that the coefficient of PR1Y is 0.022 (with a t-value of 1.66). In other words, short sellers play a weak contrarian role to short those stocks; they choose to short the stock if the recent one-year stock performance is good. Other factors of special stocks are similar to those of non-special stocks.

3.3 Changes of SIR

Boehmer, Jones and Zhang (2008) find that the average short position lasts for 37 trading days. This finding suggests that the every quarter comes with different stock borrowers and lenders. Hence, the dynamic relations between short interest and its determinants may be different from the cross-sectional relations. In addition, Table 1 shows that the ownership level of IH_QIX is higher than that of IH_TRA and IH_DED (0.32 vs 0.11 and 0.03). Therefore, IH_QIX has a stronger relationship with SIR than do IH_TRA and IH_DED, as we find. This may result from the magnitude of these variables. We use the first difference of all of the variables to examine these relationships. Given that institutional ownership is released on a quarterly basis, we compute the changes in all of the variables on a quarterly basis, and estimate the following multivariate regression.[19]

$$\Delta SIR_{i,t} = \alpha + \beta_1 \Delta RV_{i,t} + \beta_2 \Delta D\_OPT_{i,t} + \beta_3 \Delta PR1Y_{i,t} + \beta_4 \Delta IDIO_{i,t} + \gamma_1 \Delta IO\_HHI_{i,t}$$
$$+\gamma_2 \Delta IH\_DED_{i,t} + \gamma_3 \Delta IH\_QIX_{i,t} + \gamma_4 \Delta IH\_TRA_{i,t} + \gamma_5 \Delta DCBS_{i,t} + \gamma_6 \Delta UTL_{i,t}$$
$$+\varphi_1 \Delta D\_SP500_{i,t} + \varphi_2 \Delta SPREAD_{i,t} + d_{m,t} + \varepsilon_{i,t} \qquad (2)$$

Table 5 reports the results, which are similar to those in Table 3. We focus our discussion on Model 4. First, IH_QIX, IH_TRA, DCBS, and UTL remain the most important variables for determining the changes to the SIR, and all of these are supply side variables. This means that in both the level of the SIR and changes to it, supply side variables play a more important role than demand side variables. This further corroborates the evidence from the t-values. In particular, most of the demand side variables are statistically significant at the 5 percent level only, while most of the supply side variables are significant at the 1 percent level.

Second, the coefficient of ΔPR1Y becomes negative and statistically significant (-0.025 with a t-value of 2.82). This means that short sellers act as momentum investors, to short stocks after a fall in the previous quarter, and close the short position after an increase in previous quarter returns. Finally, we find that changes of IO_HHI and SPREAD play an insignificant role in determining changes to the SIR, suggesting that the quarterly changes of IO_HHI and SPREAD are too small to affect it.

[Table 5]

4. Portfolio sorting

[19] In a robustness check, we compute the monthly changes in variables and obtain similar results.

The purpose of portfolio sorting is to investigate whether the demand- or supply-side variables play the more important role in determining the SIR. Given we have four demand side variables and six supply side variables in our main regression model, i.e., Model 4 in Table 3, sorting portfolios based on each and every variable might result in some portfolios with too few observations, and results too cumbersome to interpret and read. We adopt Principal Component Analysis (PCA) to extract the demand and supply side components.[20] For the demand side, we run the PCA on the RV, D_OPT, PR1Y, and DISP to extract the 1st principal component named as DEMAND. For the supply side, we run the PCA on the IO_HHI, IH_DED, IH_QIX, IH_TRA, DCBS, and UTL to extract the 1st principal component named as SUPPLY.

We first sort the SIR by DEMAND and SUPPLY into 10 deciles independently in each month (1 is the lowest and 10 is the highest). Panel A of Table 6 reports the results. Panel A shows that when sorted by DEMAND, SIR increases from decile 1 (SIR is 0.025) to decile 8 (SIR is 0.048), then remains stable from deciles 8 to 10. The difference in SIR between deciles 10 and 1 is 0.022, which is significant at the1 percent level. The last two columns of Panel A show that when sorted by SUPPLY, the SIR increases monotonically from decile 1 (SIR is 0.014) to decile 10 (SIR 0.089). The difference in SIR between deciles 10 and 1 is 0.075, significant at 1 percent level. The difference in SIR between 10 and 1 deciles sorted by SUPPLY is more than three times that for those sorted by DEMAND. Consistent with the regression results, Panel A shows that supply side variables seem more important than demand side variables in determining the SIR.

[Table 6]

[20] We thank a referee for the suggestion of using PCA. For details of PCA results, please refer to the Internet Appendix (available on the website of *Accounting and Finance*).

Panel B provides the average SIR when sorted by DEMAND and SUPPLY independently in 5x5 groups. The results clearly show that SUPPLY is more important than DEMAND. For each DEMAND group (from d1 to d5), the difference in SUPPLY groups (s5-s1) is positive and significant. For example, for the d1 groups, the SIRs between s1 and s5 are 0.018 and 0.111. Higher supply is related to higher SIR. The DEMAND sorts are unable to absorb the SUPPLY effect on SIR. For each SUPPLY group (from s1 to s5), the difference in DEMAND groups (d5-d1) is only positive and significant for the s1 group. Indeed, the difference between d5 and d1 turns out to be negative and significant in the s2 to s5 groups, which is inconsistent with the prediction. The SUPPLY sorts, however, absorb the DEMAND effect on SIR.[21]

We further investigate the firm characteristics of subgroups s5d1 and s5d5, which show the opposite results for SIR from those predicted. Though stocks of both groups are free from short-sale constraints, the average IO, DCBS and UTL for the s5d1 group are 0.94, 1.26 and 0.32, while the corresponding numbers for the s5d5 group are 0.91, 1.04 and 0.19. We find that stocks in the s5d1 and s5d5 are very different. First, the number of observations of stocks in s5d1 group is 2953, while that in the s5d5 group is 38587. Second, most s5d1 stocks are small stocks with low stock prices and market capitalization. The average price is 21.5 in the s5d1 group, while it is 44.1 in the s5d5 group. The average market capitalization is 1286 million dollars in the s5d1 group while it is 5050 million dollars in the s5d5 group. Third, the previous one-year stock performance (PR1Y) is -0.19 for the s5d1 group and 0.42 for the s5d5 group. The results show that s5d1 group stocks are from small firms with poor recent stock performances. More importantly, investors believe such poor stock performances will continue into the future, and choose to short them. Therefore, short-sellers act like momentum investors for the s5d1 group.

[21] In a robustness check, we find the same results in sub-samples similar to those in Panel B of Table 4. Though not reported, those results are available on request.

We further confirm this conjecture by estimating a multivariate regression (Model 4 in Table 3) for the s5d1 group, and find the coefficient on PR1Y is -0.099 (with a t-value of 4.50). The corresponding coefficient for the s5d5 group is 0.03 (with a v-value of 3.02).

In summary, Table 6 further confirms the regression results, showing that the supply side variables are more important in determining short-selling activity.

5. Conclusions

Short selling plays an important role in the financial market, as it enhances the market's quality and liquidity. The implications of short selling have attracted much attention from academics, ranging from future return predictability to the real effects on corporate activity (see, e.g., Grullon, Michenaud and Weston, 2015; Chen, Zhu and Chang, 2018). However, understanding short selling itself in terms of its determinants, has received relatively less attention. This paper fills this gap by simultaneously investigating the effects of demand and supply on short selling at the firm level.

We examine a large panel of firms on the NYSE/Amex/Nasdaq from 2003 to 2015. We find that the supply side plays a more important role in determining short selling. The IO of quasi-indexer type is the most important supply side variable, while the arbitrage and hedging with options market is the most important demand side variable. We find that changes in supply side variables are more important than changes in demand side variables in determining the changes to the short interest ratio. Finally, the results from principal component analysis, and the portfolio sorting approach, confirm that the supply side is more important than the demand side in determining short selling. Our results provide a comprehensive understanding of demand for, and supply of, short selling activity.

References

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

Asquith, P., P. A. Pathak, and J. Ritter, 2005, Short interest, institutional ownership, and stock returns, *Journal of Financial Economics* 78, 20.243–276.

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

Blau, B. M., and C. Wade, 2013, Comparing the information in short sales and put options, *Review of Quantitative Finance and Accounting* 41, 567–583

Boehme, R. D., B. R. Danielsen, and S. M. Sorescu, 2006, Short sales constraints, differences of opinion, and overvaluation, *Journal of Financial and Quantitative Analysis* 41, 455–487.

Boehmer, E., C. M. Jones, and X. Zhang, 2008, Which shorts are informed?, *Journal of Finance* 63, 491–528.

Boehmer, E., C. M. Jones, and X. Zhang, 2012, What do short sellers know?, Working paper, Columbia University.

Boehmer, E., C. M. Jones, and X. Zhang, 2013, Shackling short sellers: The 2008 shorting ban, *Review of Financial Studies* 26, 1363–1400.

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

Bushee, B., 1998, The influence of institutional investors on myopic R&D investment behavior, *Accounting Review* 73, 305–333.

Bushee, B., 2001, Do institutional investors prefer near-term earnings over long-run value?, *Contemporary Accounting Research* 18, 207–246.

Carhart, M. M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Chang, E., T. Lin, and X. Ma, 2015, Does short selling discipline overinvestment?, Working paper, University of Hong Kong.

Chen, H., Y. Zhu, and L. Chang, 2018, Short-selling constraints and corporate payout policy, *Accounting & Finance*, In press.

Chen, H., H. Desai, and S. Krishnamurthy, 2013, A first look at mutual funds that use short sales, *Journal of Financial & Quantitative Analysis* 48, 761–787.

Chen, J., H. Hong, and J. C. Stein, 2002, Breadth of ownership and stock returns, *Journal of Financial Economics* 66, 171–205.

Christophe, S. E., M. G. Ferri, and J. J. Angel, 2004, Short-selling prior to earnings announcement, *Journal of Finance* 59, 1845–1875.

Cohen, L., K. B. Diether, and C. J. Malloy, 2007, Supply and demand shifts in the shorting market, *Journal of Finance* 62, 2061–2096.

Corwin, S., and P. Schultz, 2012, A simple way to estimate bid-ask spreads from daily high and low prices, *Journal of Finance* 67, 719–759.

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

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

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

Diamond, D. W., and R. E. Verrecchia, 1987, Constraints on short-selling and asset price adjustment to private information, *Journal of Financial Economics* 18, 277-312.

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

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

Engelbert, J. E., A. V. Reed, and M. C. Ringgenberg, 2012, How are shorts informed? Short sellers, news, and information processing, *Journal of Financial Economics* 105, 260–278.

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

Goldman Sachs, 2009, Hedge fund trend monitor (August 24).

Grullon, G., S. Michenaud, and J. P. Weston, 2015, The real effects of short-selling constraints, *Review of Financial Studies* 28, 1737–1767.

Grundy, B. D., B. Lim, and P. Vervijmeren, 2012, Do option markets undo restrictions on short sales? Evidence from the 2008 short-sale ban, *Journal of Financial Economics* 106, 331–348.

Hanson, S. G., and A. Sunderam, 2014, The growth and limits of arbitrage: Evidence from short interest, *Review of Financial Studies* 27, 1238–1286.

Helmes, U., J. Henker, and T. Henker, 2017, Effects of the ban on short selling on market prices and volatility, *Accounting & Finance* 57, 727–757.

Hu, Jianfeng, 2014, Does option trading convey stock price information?, *Journal of Financial Economics* 111, 625–645.

Jegadeesh, N., and S. Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Jain, C., P. Jain, and T. H. McInish, 2012, Short selling: The impact of SEC Rule 201 of 2010, *Financial Review* 47, 37–64.

Jain, A., and C. Jain, 2015, Fails-to-deliver before and after the implementation of Rule 203 and Rule 204, *Financial Review* 50, 611–636.

Jain, A., P. Jain, T. H. McInish, and M. McKenzie, 2013, Worldwide reach of short selling regulations, *Journal of Financial Economics* 109, 177–197.

Jiao, Y., M. Massa, and H. Zhang, 2016, Short selling meets hedge funds 13F: An anatomy of informed demand, *Journal of Financial Economics* 122, 544–567.

Jiang, L., and L. Pang, 2016, Does short selling enhance the allocational role of stock price? Evidence from Hong Kong, *Journal of International Financial Management & Accounting* 27, 269–297.

Kaplan, S. N., T. J. Moskowitz, and B. A. Sensoy, 2013, The effects of stock lending on security prices: An experiment, *Journal of Finance* 68, 1891–1936.

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

Kot, H. W., 2007, What determines the level of short-selling activity?, *Financial Management* 36, 123–141.

Massa, M., F. Wu, B. Zhang, and H. Zhang, 2015, Saving long-term investment from short-termism: the surprising role of short selling, Working paper, INSEAD.

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

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

Ofek, E., M. Richardson, R. F. Whitelaw, 2004, Limited arbitrage and short sales restrictions: Evidence from the options markets, *Journal of Financial Economics* 74, 305–342.

Pan, J., and A. M. Poteshman, 2006, The information in option volume for future stock prices, *Review of Financial Studies* 19, 871–908.

Petersen, M. A., 2009, Estimating standard errors in finance panel data sets: Comparing approaches, *Review of Financial Studies* 22, 435–480.

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

Stambaugh, R. F., J. Yu, and Y. Yuan, 2015, Arbitrage asymmetry and the idiosyncratic volatility puzzle, *Journal of Finance* 70, 1903–1948.

Appendix: Variable definitions

D_OPT	A dummy variable indicating the availability of stock options (1 for yes, 0 for no).
D_SP500	A dummy variable indicating inclusion in the S&P 500 Index (1 for yes, 0 for no).
DCBS	Monthly average of daily cost of borrow score, 1 for cheapest and 10 for the most expensive to borrow.
DISP	DISP is the dispersion of all analysts' forecasts of a firm's earnings for the current fiscal year (IBES fiscal year period "1") scaled by mean of forecasts (analyst coverage>=3 only).
FEE	Monthly average of daily borrowing fee.
IDIO	Idiosyncratic risk, defined as mean squared error of residuals of daily stock returns within last three month from Carhart (1997)'s 4-factor model.
IH_DED	Dedicated institutional investors (Bushee, 2001)
IH_QIX	Quasi-indexer institutional investors (Bushee, 2001)
IH_TRA	Transient institutional investors (Bushee, 2001)
IO	Institutional ownership scaled by the number of outstanding shares.
IO_HHI	IO concentration by the Hirschman-Herfindahl index
IO_TOP5	Largest five institutional shareholders scaled by IO.
ILLIQ	Ahihud's (2002) illiquidity measure.
LNSIZE	Natural logarithm of SIZE.
PR1Y	The past 1-year stock return.
RV	Relative valuation measure developed by Stambaugh, Yu and Yuan (2015). It takes a value from 1 to 10; Higher RV predicts higher return.
SIR	Short interest scaled by the number of shares outstanding.
SIZE	Market capitalization in million dollars.
SPREAD	The average daily closing bid-ask spread within 1 month.
UTL	Stock loan utilization, defined as beneficial owners (BO) on loan value divided by BO inventory value.

Figure 1: Number of firms have report short interest and short interest ratio

This figure reports the number of stocks that have reported short interest and the average of short interest ratio (SIR) on the NYSE/Amex/Nasdaq common stocks. SIR is short interest scaled by the number of shares outstanding. The sample period covers 2003.7-2015.12. The short-interest data are obtained from the Compustat.

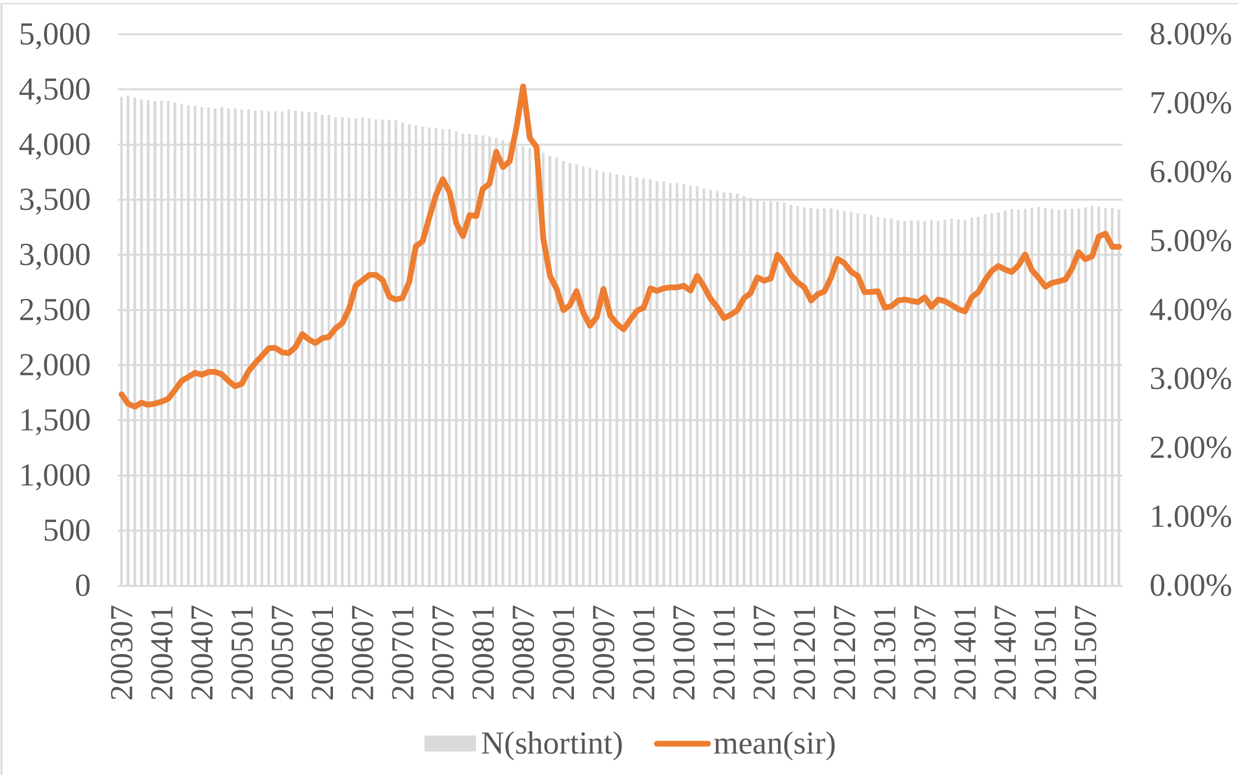

Table 1 Summary statistics

This table reports the summary statistics of variables used in the study. All variables are defined in the appendix. The sample period covers 2003.7-2015.12 for the NYSE/Amex/Nasdaq common stocks. The stock return data are obtained from CRSP. The accounting data are obtained from Compustat. The short-interest data are obtained from the Compustat. The IO data are obtained from Thomson Financial's institutional database and Bushee's website. The options availability is obtained from Option Metrics Database. The analysts' earnings forecasts data is obtained from I/B/E/S summary file. The stock loan data is obtained from Markit Equity Lending Database.

	N	mean	s.d.	min	p25	p50	p75	max
SIR	639152	0.0378	0.0539	0	0.0014	0.0192	0.0503	1
D_OPT	639152	0.5313	0.4990	0	0	1	1	1
DISP	355839	0.2021	0.5788	0	0.0181	0.0456	1335	9.3397
IDIO	634800	0.0277	0.0211	0.0041	0.0142	0.0217	0.034	0.2389
RV	638478	5.5814	1.0671	1.1	4.9	5.7	6.3	9.5
PR1Y	609784	0.1531	0.6357	-0.9654	-0.1854	0.0741	0.3468	8.6972
IH_DED	639152	0.031	0.0766	0	0	0	0.0342	1
IH_QIX	639152	0.3288	0.2637	0	0.0676	0.3123	0.5474	1
IH_TRA	639152	0.117	0.1205	0	0.0115	0.088	0.183	1
IO	639152	0.5249	0.3320	0	0.2107	0.5649	0.8274	1
IO_HHI	621569	0.1637	0.2035	0.0127	0.0445	0.0766	0.1905	1
IO_TOP5	631237	0.5805	0.2510	0	0.3772	0.5154	0.8003	1
DCBS	523170	1.5502	1.3992	1	1	1	1.1765	10
FEE	283162	0.0185	0.0671	-0.0009	0.0038	0.0038	0.0053	1.1978
UTL	574229	0.1791	0.2245	0	0.014	0.0846	0.2596	1
SIZE	635034	3683	17048	0.15	90	363	1574	750710
D_SP500	639152	0.1077	0.3100	0	0	0	0	1
ILLIQ	638304	1.1306	3.7246	0	0.0011	0.0106	0.1713	19.7147
SPREAD	639152	0.0116	0.0411	0	0.001	0.0025	0.0102	1

Table 2 Pearson's Correlations

All variables are defined in the appendix. The sample period covers 2003.7-2015.12 for the NYSE/Amex/Nasdaq common stocks. The stock return data are obtained from CRSP. The accounting data are obtained from Compustat. The short-interest data are obtained from the Compustat. The IO data are obtained from Thomson Financial’s institutional database and Bushee’s website. The options availability is obtained from Option Metrics Database. The stock loan data is obtained from Markit Equity Lending Database.

	SIR	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
01 D_OPT	0.31															
02 IDIO	0.00	-0.29														
03 RV	-0.07	0.04	-0.16													
04 PR1Y	-0.02	0.02	-0.17	0.18												
05 IH_DED	0.03	0.09	-0.07	0.02	0.00											
06 IH_QIX	0.29	0.43	-0.36	0.13	-0.01	0.03										
07 IH_TRA	0.33	0.36	-0.13	0.06	0.14	0.03	0.46									
08 IO	0.35	0.54	-0.40	0.13	0.06	0.22	0.77	0.55								
09 IO_HHI	-0.24	-0.45	0.39	-0.09	-0.11	0.07	-0.49	-0.38	-0.60							
10 IO_TOP5	-0.29	-0.61	0.47	-0.12	-0.14	0.02	-0.60	-0.50	-0.75	0.81						
11 DCBS	0.20	-0.13	0.35	-0.21	-0.11	-0.04	-0.32	-0.19	-0.38	0.32	0.35					
12 UTL	0.49	0.13	0.11	-0.16	-0.05	0.02	0.04	0.08	0.07	-0.05	-0.06	0.28				
13 LNCAP	0.12	0.59	-0.57	0.14	0.16	0.13	0.47	0.24	0.59	-0.55	-0.74	-0.30	-0.01			
14 D_SP500	-0.08	0.27	-0.25	0.15	0.00	0.04	0.20	-0.02	0.20	-0.21	-0.33	-0.13	-0.12	0.62		
15 ILLIQ	-0.16	-0.29	0.37	-0.02	-0.13	-0.05	-0.26	-0.20	-0.34	0.44	0.43	0.12	-0.11	-0.39	-0.09	
16 SPREAD	-0.20	-0.39	0.53	-0.06	-0.16	-0.07	-0.34	-0.26	-0.44	0.50	0.53	0.18	-0.10	-0.50	-0.14	0.71

Table 3 Multivariate regressions

This table reports the following multivariate regression results.

$$SIR_{i,t} = \alpha + \beta_1 RV_{i,t} + \beta_2 D\_OPT_{i,t} + \beta_3 PR1Y_{i,t} + \beta_4 IDIO_{i,t} + \gamma_1 IO\_HHI_{i,t} + \gamma_2 IH\_DED_{i,t}$$
$$+\gamma_3 IH\_QIX_{i,t} + \gamma_4 IH\_TRA_{i,t} + \gamma_5 DCBS_{i,t} + \gamma_6 UTL_{i,t} + \varphi_1 D\_SP500_{i,t} + \varphi_2 SPREAD_{i,t} + d_{m,t} + \varepsilon_{i,t}$$

All variables are defined in the appendix. The sample period and data sources are described in Table 1.

All models include stock fixed-effects and year-month dummy variables. Following Petersen (2009), standard errors are double clustered at the firm and year-month levels. We standardize all variables to have zero mean and unit standard deviation. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
RV	-0.060***		-0.032***	-0.031***
	(-10.70)		(-5.97)	(-5.91)
D_OPT	0.200***		0.081***	0.078***
	(18.57)		(9.16)	(8.95)
PR1Y	-0.006		-0.005	-0.007
	(-1.11)		(-0.87)	(-1.35)
IDIO	0.045***		0.036***	0.047***
	(5.26)		(4.95)	(5.80)
IO_HHI		-0.081***	-0.078***	-0.073***
		(-7.37)	(-6.94)	(-6.60)
IH_DED		0.040***	0.042***	0.042***
		(5.13)	(4.86)	(4.92)
IH_QIX		0.345***	0.312***	0.313***
		(20.19)	(17.08)	(17.22)
IH_TRA		0.190***	0.180***	0.177***
		(19.59)	(17.46)	(17.33)
DCBS		0.246***	0.250***	0.248***
		(14.83)	(14.69)	(14.64)
UTL		0.257***	0.260***	0.258***
		(14.03)	(13.42)	(13.38)
D_SP500				-0.059***
				(-4.21)
SPREAD				-0.071***
				(-4.73)
N	609057	513428	493002	493002
Adj.R^2	0.602	0.696	0.708	0.709

Table 4 Robustness checks of multivariate regressions

Panel A reports the following multivariate regression results.

$$SIR_{i,t} = \alpha + \beta_1 RV_{i,t} + \beta_2 D\_OPT_{i,t} + \beta_3 PR1Y_{i,t} + \beta_4 IDIO_{i,t} + \beta_5 DISP_{i,t} + \gamma_1 IO_{i,t}$$
$$+\gamma_2 IO\_HHI_{i,t} + \gamma_3 IO\_TOP5_{i,t} + \gamma_4 IH\_DED_{i,t} + \gamma_5 IH\_QIX_{i,t} + \gamma_6 IH\_TRA_{i,t} + \gamma_7 DCBS_{i,t} + \gamma_8 UTL_{i,t}$$
$$+\gamma_9 FEE_{i,t} + \varphi_1 D\_SP500_{i,t} + \varphi_2 SPREAD_{i,t} + \varphi_3 ILLIQ_{i,t} + d_{m,t} + \varepsilon_{i,t}$$

Panel B reports the following multivariate regression results.

$$SIR_{i,t} = \alpha + \beta_1 RV_{i,t} + \beta_2 D\_OPT_{i,t} + \beta_3 PR1Y_{i,t} + \beta_4 IDIO_{i,t} + \gamma_1 IO\_HHI_{i,t} + \gamma_2 IH\_DED_{i,t} + \gamma_3 IH\_QIX_{i,t}$$
$$+ \gamma_4 IH\_TRA_{i,t} + \gamma_5 DCBS_{i,t} + \gamma_6 UTL_{i,t} + \varphi_1 D\_SP500_{i,t} + \varphi_2 SPREAD_{i,t} + d_{m,t} + \varepsilon_{i,t}$$

All variables are defined in the appendix. The sample period and data sources are described in Table 1.
All models include stock fixed-effects and year-month dummy variables. Following Petersen (2009), standard errors are double clustered at the firm and year-month levels. We standardize all variables to have zero mean and unit standard deviation. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Alternative models

	(1)	(2)	(3)	(4)	(5)	(6)
RV	-0.040***	-0.027***	-0.029***	-0.031***	-0.011	-0.013*
	(-7.50)	(-4.16)	(-5.60)	(-5.94)	(-1.64)	(-1.92)
D_OPT		0.048***	0.070***	0.080***	0.019*	0.019*
		(4.97)	(8.20)	(9.10)	(1.87)	(1.81)
PR1Y	-0.012**	-0.012*	-0.012**	-0.006	-0.031***	-0.032***
	(-2.27)	(-1.79)	(-2.22)	(-1.05)	(-4.32)	(-4.32)
IDIO	0.035***	0.052***	0.050***	0.036***	0.072***	
	(4.35)	(3.99)	(6.23)	(4.74)	(5.14)	
DISP		0.013***			0.008***	0.010***
		(4.01)			(2.63)	(3.09)
IO	0.378***					
	(17.11)					
IO_HHI		-0.118***		-0.078***	-0.003	0.000
		(-4.59)		(-6.97)	(-0.12)	(0.01)
IO_TOP5			-0.123***			
			(-9.22)			
IH_DED		0.043***	0.046***	0.042***	0.026**	0.025**
		(3.93)	(5.37)	(4.90)	(2.44)	(2.32)
IH_QIX		0.322***	0.303***	0.313***	0.248***	0.244***
		(16.30)	(16.89)	(17.23)	(11.28)	(11.08)
IH_TRA		0.185***	0.162***	0.178***	0.157***	0.159***
		(16.43)	(16.01)	(17.37)	(12.24)	(12.32)
DCBS	0.239***	0.411***	0.248***	0.250***		
	(13.90)	(16.37)	(14.55)	(14.72)		
UTL	0.267***	0.310***	0.257***	0.259***	0.720***	0.725***
	(13.63)	(12.30)	(13.37)	(13.37)	(25.34)	(25.50)
FEE					0.112***	0.118***
					(4.89)	(5.07)
D_SP500	-0.060***	-0.058***	-0.064***	-0.059***	-0.040**	-0.041**
	(-4.39)	(-4.25)	(-4.54)	(-4.20)	(-2.07)	(-2.14)
SPREAD	-0.100***	-0.089**	-0.070***		-0.240***	-0.154***

	(-5.81)	(-2.16)	(-4.73)		(-3.71)	(-2.65)
ILLIQ				-0.006		
				(-1.52)		
N	501536	312023	496180	493000	177452	177457
Adj.R^2	0.691	0.724	0.709	0.708	0.818	0.817

Panel B: Sub-samples

	(1) 2003-2008	(2) 2009-2015	(3) NYSE/Amex	(4) Nasdaq	(5) Non-financial	(6) Special
RV	-0.016**	-0.017***	-0.006	-0.046***	-0.030***	-0.058***
	(-2.40)	(-3.31)	(-0.70)	(-6.97)	(-5.28)	(-4.81)
D_OPT	0.105***	0.039***	0.027**	0.107***	0.075***	0.159***
	(8.08)	(4.34)	(2.11)	(9.91)	(8.17)	(6.73)
PR1Y	0.002	-0.019***	-0.029***	0.006	-0.005	0.022*
	(0.22)	(-4.00)	(-3.62)	(0.96)	(-0.88)	(1.66)
IDIO	0.051***	0.022***	0.047***	0.041***	0.040***	0.054***
	(5.19)	(2.72)	(3.90)	(5.21)	(5.01)	(5.78)
IO_HHI	-0.122***	-0.013	-0.087***	-0.057***	-0.079***	-0.111***
	(-7.67)	(-1.14)	(-4.77)	(-4.81)	(-6.41)	(-6.37)
IH_DED	0.068***	0.038***	0.049***	0.043***	0.047***	0.021
	(5.60)	(3.71)	(3.53)	(4.18)	(5.64)	(0.67)
IH_QIX	0.478***	0.232***	0.261***	0.352***	0.321***	0.449***
	(18.22)	(14.22)	(10.71)	(16.14)	(16.79)	(9.70)
IH_TRA	0.192***	0.145***	0.132***	0.205***	0.176***	0.249***
	(13.14)	(13.60)	(8.48)	(16.24)	(16.80)	(8.74)
DCBS	0.276***	0.076***	0.268***	0.242***	0.269***	0.188***
	(14.03)	(5.69)	(10.32)	(13.06)	(14.86)	(11.58)
UTL	0.101***	0.517***	0.289***	0.231***	0.253***	0.043***
	(9.19)	(28.38)	(12.33)	(12.33)	(12.88)	(3.27)
D_SP500	-0.060***	-0.093***	-0.082***	-0.024	-0.052***	-0.014
	(-3.35)	(-3.88)	(-4.03)	(-1.59)	(-3.53)	(-0.55)
SPREAD	-0.070***	-0.002	-0.044**	-0.067***	-0.076***	-0.086***
	(-2.72)	(-0.26)	(-2.13)	(-4.03)	(-4.61)	(-3.19)
N	223240	269722	209567	283424	399406	47915
Adj.R^2	0.769	0.820	0.704	0.723	0.709	0.862

Table 5 Multivariate regressions on quarterly changes

This table reports the following multivariate regression results.

$$\Delta SIR_{i,t} = \alpha + \beta_1 \Delta RV_{i,t} + \beta_2 \Delta D\_OPT_{i,t} + \beta_3 \Delta PR1Y_{i,t} + \beta_4 \Delta IDIO_{i,t} + \gamma_1 \Delta IO\_HHI_{i,t}$$
$$+\gamma_2 \Delta IH\_DED_{i,t} + \gamma_3 \Delta IH\_QIX_{i,t} + \gamma_4 \Delta IH\_TRA_{i,t} + \gamma_5 \Delta DCBS_{i,t} + \gamma_6 \Delta UTL_{i,t} + \varphi_1 \Delta D\_SP500_{i,t}$$
$$+\varphi_2 \Delta SPREAD_{i,t} + d_{m,t} + \varepsilon_{i,t}$$

We take the quarterly difference of all variables and use the data in the end of each quarter only. All variables are defined in the appendix. The sample period and data sources are described in Table 1. All models include stock fixed-effects and year-month dummy variables. Following Petersen (2009), standard errors are double clustered at the firm and year-month levels. We standardize all quarterly changes variables to have zero mean and unit standard deviation. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
ΔRV	-0.006		-0.007*	-0.008*
	(-1.49)		(-1.91)	(-1.90)
ΔD_OPT	0.031***		0.014***	0.014***
	(6.27)		(2.83)	(2.85)
ΔPR1Y	-0.013		-0.025***	-0.025***
	(-1.50)		(-2.84)	(-2.82)
ΔIDIO	0.043***		0.020**	0.021**
	(5.59)		(2.22)	(2.62)
ΔIO_HHI		-0.006	-0.007	-0.007
		(-0.78)	(-0.82)	(-0.82)
ΔIH_DED		0.060***	0.061***	0.062***
		(4.75)	(4.29)	(4.30)
ΔIH_QIX		0.226***	0.222***	0.222***
		(13.85)	(12.70)	(12.65)
ΔIH_TRA		0.164***	0.160***	0.160***
		(10.44)	(9.89)	(9.80)
ΔDCBS		0.168***	0.168***	0.168***
		(14.60)	(13.83)	(13.84)
ΔUTL		0.179***	0.180***	0.180***
		(5.63)	(5.50)	(5.50)
ΔD_SP500				-0.009**
				(-2.31)
ΔSPREAD				-0.011
				(-0.28)
N	194769	161779	155314	155314
Adj.R^2	0.024	0.138	0.134	0.134

Table 6 Portfolio analysis

This table reports the average SIR for univariate portfolio sorting (Panel A) and two-way independent portfolio sorting. DEMAND is the 1st principal component from Principal Component Analysis (PCA) of RV, D_OPT, PR1Y, and IDIO. SUPPLY is the 1st principal component from PCA of IO_HHI, IH_DED, IH_QIX, IH_TRA, DCBS, and UTL. Panel A report the average SIR by sorting stocks by DEMAND (or SUPPLY) in each month into 10 deciles (1 for lowest and 10 for highest demand or supply). In Panel B, we first sort stocks into 5 groups by DEMAND and SUPPLY independently in each month, then compute the average SIR. All variables are defined in the appendix. The sample period and data sources are described in Table 1. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Univariate sorting

DEMAND	SIR	SUPPLY	SIR
1	0.025	1	0.014
2	0.029	2	0.019
3	0.030	3	0.026
4	0.033	4	0.036
5	0.038	5	0.042
6	0.043	6	0.047
7	0.047	7	0.048
8	0.048	8	0.051
9	0.047	9	0.058
10	0.047	10	0.089
10-1	0.022***	1-10	0.075***

Panel B: independent sorting

	s1	s2	s3	s4	s5	s5-s1
d1	0.018	0.037	0.058	0.082	0.111	0.093***
d2	0.013	0.028	0.045	0.063	0.090	0.076***
d3	0.015	0.030	0.045	0.054	0.079	0.064***
d4	0.020	0.033	0.043	0.048	0.071	0.051***
d5	0.031	0.035	0.041	0.040	0.066	0.035***
d5-d1	0.013***	-0.002**	-0.017***	-0.041***	-0.045***	