5-2019

Essays in asset pricing

Wanshan SONG
Singapore Management University, wanshansong.2014@pbs.smu.edu.sg

ESSAYS IN ASSET PRICING

WANSHAN SONG

SINGAPORE MANAGEMENT UNIVERSITY

2019

Essays in Asset Pricing

Wanshan Song

Submitted to Lee Kong Chian School of Business in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Finance

Dissertation Committee:

Ekkehart Boehmer (Chair)
Keppel Professor of Finance
Singapore Management University

Joe Zhe Zhang
Associate Professor of Finance
Singapore Management University

Jianfeng Hu
Assistant Professor of Finance
Singapore Management University

Ashish Tiwari
Professor of Finance
University of Iowa

Singapore Management University
2019

Essays in Asset Pricing

Wanshan Song

Singapore Management University, 2019

Abstract

My dissertation centers on two areas related to market microstructure. First, the role of retail traders, or the information they have, for financial markets. I use retail short selling as an example to narrow down the topic and study their trading patterns and trading strategies. Second, the role of passive indexers such as index funds or ETFs, for financial markets. I am trying to analyze the channels through which the nominally uninformed traders have influences on the return and market quality of the underlying stocks.

In Chapter 2, I study retail short sellers. It is interesting to paint retail short sellers in a positive light, because most of the literature assume that retail investors are noise traders and less likely to take short selling positions. First, by using a novel short sell transaction data from 2010 to 2016, this paper is the first to provide a comprehensive sample of short selling initiated by retail investors. I find that retail short selling is not limited, which takes up around 11% of retail trading. Second, using this sample, I find that retail short selling can predict negative stock returns. A trading strategy that mimics weekly retail shorting earns an annualized risk-adjusted value-(equal-) weighted return of 6% (12.25%). Their predictive ability is beyond that coming from overall retail investors as a group or from off-exchange institutional short sellers. Third, my results suggest that retail short sellers can profitably exploit public information, especially when it is negative. Retail short sellers also tend to be contrarians who provide liquidity when the market is one-sided due to (institutional) buying pressures. Therefore, this

paper broadens our understanding on the heterogeneity of short sellers, sheds new light on the strategies of informed traders, and complements a growing literature about the informativeness of retail investors.

In Chapter 3, I am mainly working on passive investing. Since the decision to buy or sell stocks is often directed by broader fund flows and rebalancing and not typically by stock fundamentals, we construct proxies for the two sources of trading in passive investments: one is proportionally flow-induced trading; the other is disproportionally index rebalancing. Next, we consider systematic information and provide three measures of price efficiency: price delay, variance ratio, and return synchronicity. We find that indexing significantly increases price efficiency, especially the market- (industry-) wide information. There are two related channels that drive this positive effect. First is through arbitrage: passive investing causes price discrepancies and decreases arbitrage costs, which in turn increases the speed with which systematic information is incorporated into stock prices. The second driver of the positive effect is through short selling: stocks that are added to indexes increase the available lendable shares, which reduces the cost of short selling and thus makes the incorporation of (negative) information into price faster. Overall, this paper establishes and explains the link between indexing and market quality.

In Chapter 4, I am instead starting from passive holding. Especially, index funds right now are the largest stake holders of most SP500 stocks, and they are regarded as long-term holders. It is therefore important to see how the fund managers are motived to monitor, vote, and engage with firm-level governance and long-term performance. We find that the stocks with the longest passive holding indeed outperform. We stress that the active monitoring role of passive funds contributes to long-term value creation.

Table of Contents

Acknowledgements

First, I would like to express my deepest appreciation to my supervisor Professor Ekkehart Boehmer, for his insightful guidance and continuous support of my PhD study and research. He is a respectable, intelligent, and charming professor. I will always be grateful for the supervision from him. The opportunity to learn from him is precious in my whole life.

I would also like to thank Professor Joe Zhang for his patient suggestions through my research, and my other committee members Professor Ashish Tiwari and Professor Jianfeng Hu for their valuable comments on the research work. In addition, I would like to thank Professor David Hirshleifer and Professor Chong Huang from UC Irvine for their support during my visit. They really helped me in different ways towards completion of my dissertation.

In addition, I would also like to thank my friends Xinrui Duan and Hexin Ju. Thank you for your support and company during my entire PhD life in Singapore. Your encouraging words lifted me over many difficult situations.

Lastly, I want to show my gratitude to my parents for their unconditional love, and my husband, Eric Liu, who always stands by my side and takes care of me through all the good and challenging times.

Chapter 1
Introduction

My dissertation centers on two areas related to market microstructure. First, the role of retail traders, or the information they have, for financial markets; and second, the role of passive indexers, such as index funds or ETFs, for financial markets. My goal is to better understand the nature of these traditional "uninformed" traders and to reconcile them with earlier research. Below, I will describe each of these areas in more details.

In Chapter 2, "Smart Retail Traders, Short Sellers, and Stock Returns" primarily focuses on trading patterns and trading strategies of retail short sellers. Retail investors are always regarded as noise traders and less likely to take short selling positions. Earlier literatures indicate that retail investors always underperform the market, even before considering transaction costs. Most of the papers use behavior biased stories to explain retail trading. For example, retail investors are *overconfident* in that they think they know more than they actually know. The more frequently they trade, the more underperforming their investment will be (Odean, 1999; Barber and Odean, 2000; 2001). Retail investors also exhibit *disposition* effect in that they tend to hold losing portfolios for longer time and sell winning portfolios too soon (Odean, 1998; Barber, Lee, and Odean, 2007). Retail investors are driven by *attention* and they tend to buy the stocks which are heavily discussed in the media (Barber and Odean, 2008; Da, Engelberg, and Gao, 2011). Even though retail trading is positively related to subsequent short-term returns, there will be return reversals in the long term, *i.e.* from the 5th week (Barber, Odean, and Zhu, 2008). But recent literatures find different results. They find that net retail order imbalance can positively predict short-term returns, and there don't exist long-term reversals (Kaniel, Saar, and Titman, 2008; Kaniel, Saar, Liu, and Titman, 2012; Kelley and

Tetlock, 2013, 2017; Boehmer, Jones, and Zhang, 2018). It seems that retail investors are informed traders, inconsistent with earlier literatures.

With the conflicting results, in my opinion, there are at least three issues here. First, almost all the papers use net order imbalances to measure retail trading, which means that they don't consider the offsetting buys and sells from retail trading. The large offsetting trading might explain why the traditional literature concludes that retail investors are noise traders. Second, almost all the papers suffer sample limitations. They use limited samples, either from one stock exchange or from one brokerage firm. It is hard to get the whole picture without the full sample. Third, even though they find that retail trading is informed, they don't give specific examples on the types of retail traders, or they don't consider the heterogeneity in the retail groups. Therefore, I think, it is important to start from the directional trading and solve the above three issues. That is why I want to narrow down the retail group and focus on retail short sellers.

Examining the data that are disclosed as required by Rule 606 of Regulation NMS reveals that nearly all retail orders are routed to OTC market makers. As part of the routing process, the OTC market makers provide retail orders with a small price improvement. This, in turn, allows to follow Boehmer, Jones, and Zhang (2018) and identify retail traders. Motivated by these observations, my sample starts with FINRA off-exchange short sell transactions data, which must be reported to the public since August 2009. To measure retail short selling, I select the off-exchange short sell orders that experience sub-penny price improvement compared to the NBBO. As far as I know, my paper is the first to provide a comprehensive sample of retail short selling and analyze its properties.

In my sample, for each stock-day, retail short selling covers around 10.92% of retail trading, 6.36% of off-exchange short selling, and 0.78% of total trading volumes. This suggests that retail short selling is not rare. Using this sample, I find that retail short sellers can predict negative stock returns. A portfolio that mimics weekly retail shorting earns a risk-adjusted value- (equal-) weighted return of 0.024% (0.049%) in the next 20 days, or 6% (12.25%) annualized. Their predictive ability is beyond what rises from retail investors as a group or from off-exchange institutional short sellers.

It is interesting to paint retail short sellers in a positive light, because most of the literature focuses on institutional short sellers and ignores the allegedly uninformed retail short sellers. I shed light on two hypotheses related to these issues. First, I find that retail short sellers process and act on unique information, beyond that coming from other informed traders. They benefit as this information is fully incorporated into prices. Second, when impatient institutional buyers are in the market, retail short sellers offset temporary price pressures. In this case, they may receive compensation for the provision of liquidity and benefit from the gradual reversal of price pressure. In both cases, retail traders' actions make markets more efficient. This paper broadens our understanding on the heterogeneity of short sellers, sheds new light on the strategies of informed traders, and complements a growing literature about the informativeness of retail investors.

In recent years, we can observe there is an obvious shift from active investing to passive investing. According to some estimates, by 2024, index funds will hold over 50% of the market. Given the increased pattern in passive investing, it is important to test how it affects underlying market quality. We discuss it in Chapter 3 "Passive investing, Stock Price Efficiency, and Liquidity". Since in passive investing, the decision to buy or sell stocks is often directed by broader fund flows and rebalancing and not typically by stock fundamentals, therefore, we construct proxies for the two sources of trading in passive investments: one is proportionally flow-induced trading; the other is disproportionally index rebalancing. Next, we consider systematic information and provide three measures of price efficiency: price delay, variance ratio, and return synchronicity. We find that indexing significantly increases price efficiency, especially the market- (industry-) wide information. There are two related channels that drive this positive effect. First, through arbitrage: passive investing causes price discrepancies and decreases arbitrage costs, which in turn increases the speed with which systematic information is incorporated into stock prices. The second driver of the positive effect is short selling: stocks that are added to indexes increase the available lendable shares, which reduces the cost of short selling and thus makes the incorporation of (negative) information into price faster. Overall, this paper establishes and explains the link between indexing and market quality.

In addition, William McNabb III, CEO of Vanguard, talks about index holding and states that “We want to see our clients’ investments grow over the long term, and good governance is a key to helping companies maximize their returns to shareholders.” Therefore, in Chapter 4, we test this thought from long-term passive holding perspective. In the paper “Long-term Index Fund Ownership and Stock Returns”, we begin with the large indexers and study whether indexers’ efforts to improve stock returns are effective. We find that the stocks with the longest passive duration or the smallest passive churn ratio indeed outperform. We address the endogeneity problem by exploiting the annual Russell index reconstructions. In additional test, we explicitly exclude the effect coming from active funds or closet indexers so that out results are not due to mis-classified active funds. To conclude, we stress that the active monitoring role of passive funds contributes to long-term value creations.

Chapter 2

Smart Retail Traders, Short Sellers, and Stock Returns

Abstract

Using a novel short sell transaction data from 2010 to 2016, this paper is the first to provide a comprehensive sample of short selling initiated by retail investors. I find that retail short selling can predict negative stock returns. A trading strategy that mimics weekly retail shorting earns an annualized risk-adjusted value-(equal-) weighted return of 6% (12.25%). Their predictive ability is beyond that coming from overall retail investors as a group or from off-exchange institutional short sellers. My results suggest that retail short sellers can profitably exploit public information, especially when it is negative. Retail short sellers also tend to be contrarians who provide liquidity when the market is one-sided due to (institutional) buying pressures.

2.1 Introduction

Academics regard short sellers as informed traders who help correct short-term deviation of stock prices from fundamental values. Boehmer, Jones, and Zhang (2008) attribute most of the price discovery that accrues with short selling to institutional traders and provide limited evidence about retail short sellers. Yet because of data limitations, the impact of retail traders including retail short sellers on stock pricing still remains unsettled. Earlier literature suggests that retail investors are noise traders and do not contribute to price discovery (Odean, 1998; Barber and Odean, 2000; 2001;2008), but recent literature disagrees with this view (Kaniel, Saar, and Titman, 2008; Kaniel, Liu, Saar, and Titman, 2012; Kelley and Tetlock, 2013; Barrot, Kaniel, and Sraer, 2016; Boehmer, Jones, and Zhang, 2017). Here, I examine whether retail traders' short selling can account for the information in retail trades.

A review of the order routing disclosures required by Rule 606 of Regulation NMS reveals that nearly all retail orders are routed to OTC market makers[1]. As a part of routing process, the OTC market makers always provide retail orders with a small price improvement. Exploiting this observation, Boehmer, Jones, and Zhang (2017) identify retail trades and find that they are quite informed, but the precise nature of retail investors' information remains unclear. Moreover, the heterogeneity in the retail investors' group, such as *retail short sellers*, is not clear either.

In this paper, I dissect the joint presence of short sellers, retail traders, and other over-the-counter trading (that we mostly attribute to dark pool) by exploiting a novel comprehensive short sell transaction level dataset. I also ask which traders' order flows are related to liquidity provision and which are related to price discovery. Overall, I provide new evidence by focusing on the return predictability and trading strategy of directional retail traders: retail short sellers.

[1] See "SEC 2010 Concept Release on Equity Market Structure".

Additionally, I highlight the heterogeneity among retail investors. I can also address the effect of trading platforms and the degree of immediacy that either side of trade demands. My results are consistent with the information hypothesis that retail short sellers have and act on unique information beyond that coming from other investors, but they are also consistent with strategic liquidity provision that retail short sellers provide in one-sided markets.

I construct a comprehensive sample and provide extensive evidence of retail short selling. My sample starts with FINRA TRF short sale transaction level data, which must be reported to the public since August 2009. This data set provides detailed observations on the OTC short sales (i.e., short sales not reported to an exchange), and is a subset of the transactions reported in the daily TAQ (exchange code "D"). Exploiting the availability of short sell transaction data provided by FINRA, I identify marketable retail short sell orders as those short sale orders that experience sub-penny price improvement compared to the national best bid and offer (Boehmer, et al., 2017). My sample is from January 2010 to December 2016 and includes total 5,332 stocks, with an average of 3,532 stocks each day. On average, retail short sellers trade 7,925 shares per stock-day, which is around 10.92% of retail trading, 6.36% of off-exchange short selling, and about 0.78% of total trading volumes.

I find that retail short sales significantly predict negative future returns. A portfolio that mimics weekly retail shorting earns a risk-adjusted value-(equal-) weighted return of 0.024% (0.049%) during the next 20 trading days, or 6% (12.25%) annualized. Using a cross-sectional regression, I find that most of the predictive power survives after controlling firm characteristics, such as firm size, book to market ratio, turnover, volatility, and past returns. The predictive power also remains significant after controlling for the trading activity of other presumably informed traders. Specifically, I control for the order imbalance of retail investors, defined as retail buy volumes minus retail long sell volumes. I also control for other off-exchange short sales, which mostly reflect shorting in dark pools by non-retail traders.

The results suggest that retail short sellers are informed traders. But are there any differences between retail investors’ overall trades and their short sales? Retail investors who are able to sell short stocks may be different from the typical retail investors, who don’t necessarily have margin accounts that allow short selling (Gamble and Xu, 2017). To address this question, I conduct double sorts. The first sort is on retail trading and the second sort is on retail shorting. This approach provides portfolios that vary in terms of retail shorting activity but hold constant the level of overall (other) retail trading.

I find that retail short selling can predict future negative returns in all the five quintile portfolios, suggesting that retail short sellers have additional information over the information that retail investors have in general. In addition, I use cross-sectional regressions to test whether retail short sellers trade on the same particular types of firms as other retail traders do. The results show that retail short sellers can predict returns in all market cap and all liquidity groups. Their predictions are stronger for liquid stocks. In contrast, other retail traders can better predict small-cap and illiquid stocks, while they lose predictive power for returns of large-cap and liquid stocks. This evidence suggests that, different from the typical retail investors who have an information advantage in small firms, retail short sellers are informed about both large and small stocks.

How should we interpret the fact that retail short sellers as a group seem to be able to predict stock returns? I shed light on two stories about retail shorting and stock prices. The first story holds that retail short sellers are informed traders who are at least as sophisticated as institutional short sellers. In this view, retail short sellers have access to public information that they can exploit in their trading. They benefit as this information is fully incorporated into stock prices (Christophe, Ferri, and Angel, 2004; Engelberg, Reed, and Ringgenberg, 2012; Boehmer, Jones, and Zhang, 2015).

To test this hypothesis, I select a sample of news related to corporate earnings' information and analysts' information. By separating days when there is news from those days when there is no news, I find that retail short sellers' predictive power for next week's return doubles on news days compared to non-news days. Next, I select the days with only negative news announcement, including the days with negative unexpected earnings' surprises, analyst recommendation downgrades, and negative analyst earnings revision. I find that retail short sellers have much larger predictions for both next week's returns and next month's returns on negative news days. Specifically, the prediction of retail short selling on negative news days is -0.074%, which is up to 5 times more informative about the cross-section of next week's return compared to other days. The prediction for next month's return on negative news days is -0.119%, which doubles compared with other non-negative news days. Put differently, negative news announcement covers around 16.3% of all the underperformance related with retail shorting. Finally, I test the differences in information advantage among the group of retail investors. There is no evidence that other retail traders (other than retail short sellers) act on such public information.

Aside from an information advantage, retail short selling has another potential explanation. The second hypothesis is related to liquidity provision. According to this story, retail short sellers step in and trade against buying pressure in the market when buyers are more aggressive. As the buying pressure subsides, prices revert to fundamental values and short sellers cover their positions and earn negative (*i.e.,* profitable) returns. In this view, retail short sellers' trading patterns and their predictive power result from gradual reversals of price pressures (Nagel, 2012; Diether, et al., 2009; Comerton-Forde, et al., 2016).

To test whether liquidity provision is consistent with the data, I first look at the sensitivity of retail short sells to past returns and find that retail short sellers are strong contrarians. High retail shorting follows positive past returns, suggesting that retail short sellers step in to initiate or increase their shorting positions after prices rise. This is broadly consistent with liquidity provision. Second, I directly look at the relation between retail short sellers and contemporaneous and past buying pressure. When there is greater overall buying pressure (defined as a positive Lee and Ready (1991) order imbalance), retail short sellers increase their positions. By dividing overall buying pressure into retail buying pressure (defined as a positive retail order imbalance, excluding retail short sells) and institutional buying pressure (defined as a positive Lee and Ready order imbalance, but excluding the retail order imbalance), retail short sellers increase their activities more with contemporaneous institutional buying pressure. Both retail short sellers and institutional buying pressures are related with subsequent negative stock returns. This observation is also consistent with the liquidity provision story: Retail short sellers step in to provide liquidity to impatient (institutional) buyers when there is temporary contemporaneous price pressure.

In the last part of this paper, I will address the question: when do retail short sell orders relate to new information, and when do they relate to liquidity provision? I isolate the relation of informed retail short selling to returns on news days from the relation on non-news days. On non-news days, return reversals are known to be larger than on news days (Tetlock, 2010). Results indicate that retail short sell orders are less contrarian relative to past returns on days with news releases. Next, I separate the relation between retail short selling and returns on the stocks with positive buying pressures from stocks without buying pressures. Results show that retail short sell orders are more contrarian relative to past returns on stocks with positive buying pressure. Taken together, retail short sellers provide liquidity in response to liquidity shocks in the absence of news releases and in the presence of price pressure in the market.

Closely related papers that discuss retail short selling are Boehmer, Jones, and Zhang (2008) and Kelley and Tetlock (2017). BJZ (2008) study retail short sales that are executed on the New York Stock Exchange. But for retail orders, brokers tend to internalize their orders or route unwanted orders to wholesales (Boehmer, Jones, and Zhang, 2017). Brokers are not likely to route retail orders to registered exchanges. In addition, BJZ (2008) shows that short selling prediction mainly comes from institutions, while the return prediction from retail short sellers is insignificant even with the correct sign.

Similar to this paper, KT (2017) also show that retail short sales can predict negative stock returns. But compared with their study, my paper differs in three important ways. First, KT's proprietary data stem from one OTC market maker, which covers around 1/3 of all the retail short sells from 2003 to 2007. In their sample, retail short sells are around 0.13% of total shares traded for each stock-day. In contrast, my sample uses publicly available short sell transaction data and covers the most recent period from 2010 to 2016. In my sample, retail short sells are around 0.78% of total shares traded. Second, my sample allows further analysis of the differences between retail shorting and OTC shorting. This highlights the differences between retail short sales and institutional short sales executed in dark pools. This also helps to address the effect of trading platform and the degree of immediacy that either side of the trade demands. I find that retail short sellers are more aggressive than institutional dark pool short sellers. Institutional investors are more likely to route their orders to exchanges if they are more aggressive and more informative (Zhu, 2014). Third, I find that retail short sellers provide liquidity, earning compensation for doing so, and especially in one-sided markets. Their trading benefits by trading on the other side of institutional buying pressures, rather than exploiting retail net buying as indicted by KT (2017).

Overall, this paper is the first to use a publicly available dataset to study the group of retail short sellers. Relying on a comprehensive short selling dataset, I analyze marketable short sell orders from retail traders in the U.S. equity market between 2010 and 2016. I find that retail short sells are integral to both price discovery and liquidity provision. They have distinct predictive powers beyond that coming from other informed traders in general.

This paper is organized as follows. Section 2 describes the data sources and sample construction. Section 3 contains portfolio results, including a comparison between retail short selling and other retail trading. Section 4 presents the main cross-sectional regressions in which I use retail short sales to predict returns. Section 5 provides analysis to disentangle the information advantage hypothesis from the liquidity provision hypothesis. Section 6 concludes.

2.2 Data

2.2.1 Sample Construction

I obtain retail short sell transactions level data from two sources: retail transactions data from the daily Trade and Quote (TAQ) database, and short sell transactions data from the Financial Industry Regulatory Authority (FINRA).

I follow Boehmer et al. (2017) to identify retail transactions in the data. In the U.S., most equity trades initiated by retail investors do not take place on registered exchanges. Instead, they are executed either by wholesalers (e.g., Citadel) or via brokers internalization (e.g., Charles Schwab). Orders executed in one of these ways are reported to FINRA's Trade Reporting Facility (TRF), which will show up in TAQ with exchange code "D". In addition, as a part of routing process, brokers or wholesalers will generally provide retail orders with sub-penny price improvement over a round penny. Therefore, retail sales are filled with prices just above a round penny, and retail purchases are filled with prices just below a round penny. In contrast, institutional orders that are sent to exchanges or dark pools are prohibited from using fractional pennies. There is one exception that allows institutional participants to print crossing network execution as fractional pennies (i. e., 0.5 pennies). Conservatively, we do not label 0.4 and 0.6 as retail trades, because some crossing networks allow prints on these fractions. Therefore, I select the trades in TAQ data with exchange code "D" that additionally receive sub-penny price improvement. For $P_{i,t}$, $Z_{i,t} = 100 * mod(P_{i,t}, 0.01)$. If $Z_{i,t}$ is in the interval (0, 0.4), we mark it as a retail seller-initiated transaction. If $Z_{i,t}$ is in the interval (0.6, 1), it is a retail buyer-initiated transaction.

The second data source, available from FINRA, makes off-exchange short sell transactions publicly available since August 2009[2]. To increase the transparency of short sale transactions, after financial crisis, the Securities and Exchange Commission (SEC) has required self-regulatory organizations (SROs) to provide short sell transactions on their website. It includes the FINRA Trade Reporting Facility (TRF)[3], which provides a mechanism for the reporting of short sell transactions that take place OTC rather than on an exchange. The FINRA website provides detailed off-exchange short sale transactions from the 'consolidated tape'. The data cover the period from August 2009 to 2018 and include a market center identifier (Nasdaq TRF, NYSE TRF or ADF)[4], stock symbol, date, time, price, volume, and trader type (exempt or nonexempt from short sell rule). The FINRA short sale data are a sub-sample of off-exchange transaction data reported in the daily TAQ (exchange code "D"). This paper is the first to exploit the high-frequency FINRA TRF short selling dataset. In my sample, it covers a long sample period of 7 years from 2010 to 2016 for all national market system stocks.

In sum, exploiting the availability of short sale trade data from FINRA, I can identify retail short sell transactions. I start with FINRA's off-exchange short sale transaction prices $P_{i,t}$, and mark the short sale transactions prices above a round penny. These deviations indicate price improvement and thus a retail short sale. In fractions of one penny $Z_{i,t} = 100 *$

---

[2] Finra's website includes two types of data: The first is daily file that includes aggregate daily short volume for each firm. The second is monthly file that includes transaction-by-transaction detail of short sale trades reported to the consolidated tape. To look at transaction price of each short sell transaction that happened off-exchange, I use the monthly file.

[3] TRFs are operated by a registered national exchange (e. g., NASDAQ or NYSE). All TRFs are subject to FINRA's oversight.

[4] In my sample, I only include market center of Nasdaq TRF ('Q'), which corresponds to daily TAQ data with exchange code 'D' and TRF 'Q/T', and NYSE TRF ('N'), which corresponds to daily TAQ data with exchange code 'D' and TRF 'N.'

$mod(P_{i,t}, 0.01)$, that is 0< $Z_{i,t}$ <0.4. This approach allows me to identify the short sell transactions initiated by retail investors.

Specifically, I only include short sales during normal market hours (09:30:00 to 16:00:00) and only include shorts that are not exempt. To be included in the sample, I require matching data in daily TAQ and the FINRA data. If a stock has no retail short selling on that day, it will be noted as zero shares traded with zero number of trades. For the purpose of the analysis, within a given trading day, we sum up all the retail short sell shares or the number of retail short selling trades for a stock. Then this data is merged with stock return data from CRSP and accounting data from Compustat. I only include common stocks with share code 10 or 11 listed on one of the three main exchanges NYSE, NASDAQ, or Amex with a minimum stock price of $1 at the previous month-end. My sample is from January 2010 to December 2016. It includes 5,332 distinct stocks totally with an average of 3,532 stocks on each trading day.

2.2.2 Summary Statistics

Table 1 provides summary statistics of the variables used in the paper. I pool observations across stocks and across days, and calculate the mean, standard deviation, median, 25th, and 75th percentile. There are total 6,224,866 observations.

I use the millisecond data from daily TAQ and calculate shares traded and the number of trades for each stock-day from 2010 to 2016. I calculate trading volume during normal market hours and filter out trades with prices below zero and with a correction indicator other than zero. On average, for each stock-day, the mean trading volume (*vol*) is around 1.148 million and the mean number of trades (*trd*) is around 5,603. Next, I follow Lee and Ready (1991) to identify the transactions initiated by buyers or sellers. On average, buyer-initiated trading volume (*vol_buy*) is around 0.569 million and seller-initiated trading volume (*vol_sell*) is around 0.578 million. The number of buyer-initiated trades (*trd_buy*) and seller-initiated trades (*trd_sell*) is similar, which is around 2,800 on each stock-day.

Next, I select the OTC trades that took place away from the exchanges. These OTC trades have exchange code "D" in daily TAQ. On average, for each stock-day, OTC trading volume (*otcvol*) is around 0.394 million, which is around 1/3 of all the shares traded. I select retail marketable orders, which are the orders receiving sub-penny price improvements relative to the NBBO. If there is no retail trading on that day, both retail buys and retail sells are documented as zero on that day. On average, daily buy volumes from retail investors (*rtlvol_buy*) are 42,526 shares and the daily sell volumes from retail investors (*rtlvol_sell*) are 42,177 shares. Thus, trades from retail investors are around 7.5% of the total shares traded.

Moreover, I select OTC short sells reported on the FINRA website, and only include them in the sample when they can be matched with the daily TAQ data. If there is no OTC shorting on that day, it will be noted as zero. The average daily short sell volume that took place off-exchange (*otcshort_vol*) is 160,189 shares and the average daily off-exchange short sell trades (*otcshort_trd*) are 583 trades. Thus, OTC short selling is around 37.93% of the total OTC trading volume or 12.69% of the average total shares traded each day.

To select retail shorting, I look at OTC short sell transaction price $P_{i,t}$ and require a transaction price above a round penny. That is, let $Z_{i,t} = 100 * mod(P_{i,t}, 0.01)$ be the fraction of a penny associated with OTC short sell transaction price $P_{i,t}$, and select it if 0< $Z_{i,t}$<0.4. For each stock-day, retail short sale volume (*rtlshort_vol*) is 7,925 shares, and retail short sell trades (*rtlshort_trd*) are 24.

Finally, to evaluate the relative importance of retail short sales as a fraction of total retail trading and total off-exchange short selling, I report the relevant percentages at the bottom of Table 1. On average, retail short sale shares are around 10.92% of retail investors' traded shares (*rtlshort_rtlvol*) and around 6.36% of all OTC short selling shares (*rtlshort_otcshortvol*). Overall, retail short sell volumes are around 0.78% of all the shares traded each day (*rtlshort*), and retail short sell trades are around 0.53% of all the trades (*rtlshort_t*).

2.3 Portfolios Returns

2.3.1 Single Sort Portfolios: Retail Short Selling

To study whether retail short sells can predict stock returns, my analysis begins by constructing calendar-time portfolios where returns represent the performance of stocks with different levels of retail short selling. If retail short sellers are informed, the stocks heavily shorted by retail investors should underperform the stocks not shorted or lightly shorted by retail investors. To minimize the impact of microstructure noise on my results, I focus on weekly horizons of retail short sells, i. e., retail short sell activities over the previous 5 trading days [t-5, t-1].

Each day, I sort all the stocks into quintiles based on weekly retail short selling, where portfolio 1 consists of stocks with the least retail short selling and portfolio 5 consists of the stocks with the most retail short selling. I use 4 measures to proxy for retail short selling. The first two are unstandardized shorting measures: weekly retail short selling volumes (*rtlshort_vol_w*) and weekly retail short selling trades (*rtlshort_trd_w*); and the next two are standardized measures: weekly percentage of retail short selling volume (*rtlshort_w*), scaled by total traded shares, and weekly percentage of retail short selling trades (*rtlshort_t_w*), scaled by total number of trades. After stocks are sorted into quintiles each day, I skip one day (t=0) and then hold a value-(equal-) weighted portfolio for 20 trading days [t+1, t+20]. This process is repeated each trading day, so there are 20 overlapping portfolios. To deal with the overlap, a calendar-time approach is used to calculate the average daily returns (Jegadeesh and Titman, 1993). Thus, each trading day's portfolio return is the simple average of 20 different daily portfolio returns, and 1/20 of portfolios is rebalanced each day (Boehmer, et al, 2008).

Average daily returns for each portfolio are either value-weighted or equal-weighted across all the stocks in the portfolio. Value-weighted portfolios use market-value weights at the previous month-end, with the weights summing to one for each portfolio on each trading day. Equal-weighted portfolios are equal-weighted across portfolios on each trading day. Table 2 reports both raw returns and alphas for each portfolio. Each portfolio's alpha is the intercept from a time-series regression of its daily excess returns on the three Fama and French (1993) daily return factors, based on market, size, and book-to-market. I calculate the spread portfolio return, which is the return of the heavily shorted stocks (quintile 5) minus the return of the lighted shorted stocks (quintile 1).

The result in Table 2 suggests that retail shorting predicts negative returns, no matter what measure is used. Daily (annualized) value-weighted alphas of the spread portfolios are -0.024%, -0.027%, -0.024%, -0.027% (-6%, -6.75%, -6%, -6.75%), respectively. The significance of the spread returns is 5% or better in all regressions. Daily (annualized) equal-weighted alphas of the spread portfolios are larger, which can arrive at -0.075%, -0.074%, -0.049%, -0.045% (-18.75%, -18.5%, -12.25%, -11.25%). Both magnitude and significance are larger, compared to the value-weighted portfolio results, implying that retail short sellers are better able to predict returns for the small-cap stocks. Below, I will focus on the standardized shares measure (weekly retail short's percentage of volume *rtlshort_w*) and on the value-weighted portfolio returns.

Table 2 reports returns on the monthly horizons, which are likely to match the short sellers' horizons. In addition, I also test the return predictions at other time horizons, reported in Appendix A1. I look at the next 5 trading days' returns [t+1, t+5], the next 1 month returns [t+1, t+20], the next 2 month returns [t+21, t+40], and the next 3 month returns [t+41, t+60]. For example, using weekly retail short's share of volume (*rtlshort_w*) to divide the sample, the daily value-weighted alpha of the spread portfolio is larger in the short term [t+1, t+5], with -0.028% and a t-value of -2.695. In the longer term at [t+21, t+40], the return difference is smaller and becomes less significant, with -0.017% and a t-value of -1.783. Even in [t+41, t+60], we do not observe no return reversals, but the return difference becomes insignificant.

Since quintile 1 includes the stocks with low retail shorting and no retail shorting at the same time, I perform an additional estimation. I put the stock-days without retail shorting in the 1st quintile, and then equally divide the rest into 4 quartiles.[5] This way, return spreads will be less noisy. For example, using weekly retail short's share of volumes (*rtlshort_w*) to sort the sample, the daily value-weighted alpha of the spread portfolio during the next 20 trading days is -0.054% with a t-value of -5.119, annualized at -13.5%. This suggests that stocks heavily shorted by retail investors will underperform the stocks not shorted by retail investors.

2.3.2 Double Sort Portfolios: Retail Trading and Retail Short Selling

Table 2 suggests that retail short selling can significantly predict negative returns. But recent studies by Boehmer, Jones, and Zhang (2017), Kelley and Tetlock (2013), and Barrot, Kaniel, and Sraer (2016) find that retail investors are informed in a more general sense (*i. e.*, even when they are not shorting). This retail flow can predict future stock returns even when I do not condition on retail short sales. Because our study is close to Boehmer, *et al.* (2017), it is useful to establish whether or not retail traders' short selling contains additional information over the information that retail trades have in general. Specifically, I ask what is the difference between retail investors' long trades and their short sales? To answer this question and to

[5] Portfolio returns using this approach have one drawback: the number of stocks in each quintile is not equal. On average, there are 383 stocks each day in the 1st quintile, and 787 stocks each day in the next 4 quartiles.

highlight the differences, if any, I conduct double sorts: all the stocks are first sorted on retail trading and second sorted on retail shorting. This approach considers the predictive power of retail short sellers beyond the predictive power that retail flows have in the aggregate.

Specifically, on each trading day, I first sort all the stocks into quintiles based on weekly retail order imbalances. Within equal quintile, I then sort stocks into quintiles based on weekly retail shorting. Retail order imbalance is calculated as retail buyer-initiated volume (*rtlvol_buy*) minus retail long seller-initiated volume (*rtlvol_sell-rtlshort_vol*), scaled by total shares of volume (*vol*) for each stock-day. This approach excludes retail short sales from retail order imbalances. I skip one day and calculate the value-weighted portfolio returns using a 20-day holding period, where returns are presented in percent. As before, I use a calendar-time approach to calculate returns and adjust overlaps. Therefore, the result in Table 3 is a set of stocks that differ in retail shorting but have similar overall retail trading.

In Table 3, after controlling for weekly retail order imbalances, the predictive power of retail shorting using raw returns is preserved across the medium three retail order imbalances quintiles, and the predictive power of retail shorting using risk-adjusted returns is preserved across all the five retail order imbalance quintiles. Specifically, Column (1) presents a portfolio of stocks with the lowest retail order imbalances and column (5) presents a portfolio of stocks with the highest retail order imbalances. Comparing column (1) and column (5), stocks that are heavily bought by retail investors outperform stocks that are heavily sold by retail investors, which is consistent with Boehmer et al. (2017). Within each column, it includes stocks with different levels of retail shorting, where row (1) has the least retail shorting and row (5) has the most retail shorting. The return difference between row (5) and row (1) is smaller in the first three columns and becomes larger in the last two columns. For example, when retail buying is ranked in the 2nd column, the raw return spread and alpha spread are -0.017% and -0.022% (-

4.25% and -5.5% annualized). When retail buying is ranked in the 5th column, the raw return spread and alpha spread are -0.024% and -0.050% (-6% and -12.5% annualized).

Overall, Table 3 shows that among the stocks that have similar retail flows, the stocks where retail traders short more underperform. In addition, retail shorting is more predictive of returns within the subset of stocks that other retail investors have heavily bought[6]. Therefore, the double sort results suggest that retail shorting is based on the information that is distinct from the information that retail traders have as a group.

2.4 Regression Results

2.4.1. Return Predictions of Retail Short Selling

Portfolio results in Table 2 and Table 3 are based on univariate sorts that could reflect omitted variable bias, or other variables that predict returns could be correlated with retail shorting. In this section, I ask whether retail short sales can predict returns after controlling for trading by other informed investors and for firm characteristics.

To address this question, I use a Fama-Macbeth (1973) regression of future returns on weekly retail shorting. Analogously to the portfolio returns above, dependent variables are the next 20 days' returns [t+1, t+20]. Since I use overlapping daily data for short sells and return measures, the standard deviations of the time series are adjusted using Newey-West (1987) with 20 lags. In panel A, dependent variable is cumulative raw returns from t+1 to t+20. In panel B, dependent variable is cumulative risk-adjusted returns from t+1 to t+20, with betas estimated from previous year. All the returns are in percent.

[6] I also switch the sorting sequence. Retail net buying can only predict returns for stocks which are less likely to be shorted by retail investors. It means that, retail short selling can take advantage of other retail trading, while other retail trading cannot take advantage of retail short selling.

As control variables, I include past returns over three different horizons: the previous week [t-5, t-1] *ret(w-1)*, the previous month [t-26, t-6] *ret(m-1)*, and the previous year [t-251, t-26] *ret(y-1)*. It is important to include past returns since retail short sellers may be contrarian traders. I also include log market cap, log book to market ratio, turnover, and daily return volatility, all from the previous month. These firm characteristics are related to public information and impact future returns. Finally, I also consider trading by other investors. I include other retail investors' net buying *rtloib* (*i. e.*, net buying by retail traders who are not shorting in a particular trade), since retail trading is known to be related to returns. Here, "other" retail activity is measured as: *(rtlvol_buy-(rtlvol_sell-rtlshort_vol))/vol*. I also include non-retail off-exchange shorts, *Darkshort*. This measure is calculated as: *(otcshort_vol-rtlshort_vol)/vol*. These non-retail OTC shorts generally represent institutional shorting in dark pools. To be consistent with the previous sorts in Table 2 and Table 3, I use weekly retail order imbalance and weekly dark pool short sells. All the independent variables are standardized each day to facilitate comparisons of the coefficients. At the first stage, for each day, I estimate the following cross-sectional regression:

$$r_{i,t+1,t+20} = b_0 + b_1 * rtlshort\_w_{i,t-5,t-1} + r * x_{i,t-1} + e_{i,t} \tag{1}$$

In the second stage, I calculate time-series averages of daily regression coefficients.

Table 4 reports regression estimates. The 1st column shows the univariate impact of retail short sells. The standardized coefficient is -0.114%, with a t-statistic of 4.04, suggesting that if retail investors short sell more in a given week, the return on that stock for next month is significantly lower. In terms of magnitude, the quintile range for *rlshort_w* is (0.8522-(-0.6984))=1.5506. Multiplying the quintile difference with the standardized coefficient: -0.114%*1.5506=-0.177% (or annualized at -2.12%). The 2nd column in Table 4 adds control variables for public information. The magnitude of the retail shorting coefficient declines slightly to -0.065%, but it remains significant at the 1% level. The slightly lower coefficients suggest that retail short sellers trade on some public signals, but they have additional information that buyers or long sellers don't have. The negative coefficients on the past week's returns and the positive coefficients on the past year's returns reflect the presence of short-term return reversals at the 1% level and a marginally significant indication of long-term return momentum at the 5% level.

The 3rd column introduces non-retail dark pool short sales, *darkshort_w*. It has a coefficient of -0.085%, which is also a strong predictor of the future negative returns. For retail short selling *rtlshort_w*, the coefficients decrease from column (2) to column (3), implying that retail short sellers and dark pool short sellers partially share the same information. But even after controlling for the dark pool shorting, retail shorting can still predict negative returns with a magnitude of -0.051% and a t-statistics of 2.58. The 4th column introduces retail net buying *rtloib_w*, which can positively predict future returns consistent with the results in Boehmer et al. (2017). It has a standardized coefficient of 0.126%. For retail short selling *rtlshort_w*, the coefficients increase from column (2) to column (4), implying that the information of retail short sellers is partly different from the information that other retail traders have as a group, even when controlling for firm characteristics.

In panel B, I change dependent variables from cumulative raw returns to cumulative risk-adjusted returns. The predictive power of retail short selling is similar in both magnitude and significance and it remains a significant predictor of future negative returns. For example, in column (2), the standardized coefficient on *rtlshort_w* is -0.056% and it is significantly below 1%. The coefficient on *rtlshort_w* decreases from column (2) to column (3) and increases from column (2) to column (4), suggesting again there exists overlap of information among short sellers and distinction of information among retail investors. In addition, size becomes negatively significant and book-to-market ratio becomes positively significant.[7]

As a robustness check, I change dependent variables from next 20 days' returns to next 5 days' returns [t+1, t+5], reported in Appendix A2. Retail shorting is a strong predictor for next week's returns. When included alone as an independent variable in column (1), the standardized coefficient on retail short sells is -0.031% with a t-value of 4.13. That means the monthly predictions will be -0.124% (-0.031%*4), larger than the prediction magnitude for next month's returns of -0.114% reported in Table 4. What's more, coefficient on *darkshort_w* becomes insignificant in column (3), suggesting that dark pool shorting is less aggressive than retail shorting in the short term. Retail trading *rtloib_w* in column (4) is still a strong positive predictor for next week returns with a coefficient of 0.050%. Consistent with Zhu (2014), informed traders (*i.e.,* short sellers) have a disincentive to route the informed orders to dark venues due to more execution risk.

To sum up, Table 4 shows that short sales initiated by retail investors negatively predict future stock returns. Their predictive power is not subsumed by firm characteristics or other trader types. Two related studies by BJZ (2008) and KT (2017) also find that retail shorting can predict stock returns, but both samples are quite limited. My results hold even when controlling for the imbalances of other traders, and my results are distinct from the results for all retail activities as a group (as in BJZ, 2017).

2.4.2 Comparing Return Predictions for Retail Short Selling vs. Other Retail Trading

I have shown that retail shorting can predict stock returns. But are short sellers special, or do they trade on the same information as other retail traders do? The results in Table 4

[7] Since Panel A and B have similar results, I will focus on the results using raw returns as dependent variables. The results using risk-adjusted cumulative returns as dependent variables are qualitatively identical and available upon request.

suggest that retail order imbalances have greater predictive power than do retail short sales. To compare retail short sellers to the typical non-short retail investors, I use sub-sample cross-sectional regressions.

In table 5, all stocks are sorted into three groups based on firm size at the end of the previous month in Panel A and on firm liquidity at the end of the previous month in Panel B. Then, I re-estimate Fama-Macbeth (1973) regressions within each group. That is, all the coefficients are allowed to be different within each size group/each liquidity group. Next, I calculate the total retail order imbalance *rtloib_total_w*, which includes retail short sells. That is: *(rtlvol_buy-rtlvol_sell)/vol*. I further divide it into two parts: one is retail short sells *rtlshort_w*, calculated as *rtlshort_vol/vol*; the other is retail net buying *rtloib_w*, calculated as *(rtlvol_buy-(rtlvol_sell-rtlshort_vol))/vol*, this time excluding retail short sells.

In Panel A Table 5, Columns (1) and (2) report results for small-cap stocks. In column (1), total retail order imbalance is a strong predictor of future stock return. The standardized coefficient is 0.189% with a t-statistics of 10.18. In column (2), I divide total retail order imbalance into two parts: retail short selling and retail net buying. The coefficient on *rtlshort_w* is -0.065% and the coefficient on *rtloib_w* is 0.182%. It means for small-cap stocks, retail trading tends to have greater predictive power than retail shorting, suggesting that retail investors are better informed in small-cap stocks. Next, for medium-cap stocks in columns (3) and (4), the relative predictions of retail shorting and retail net buying are similar. However, the picture is totally different in large-cap stocks in columns (5) and (6). Only retail shorting has a significant impact on stock pricing. In column (5), the coefficient of *rtloib_total_w* is not significant any more. In column (6), retail short sales remain a strong significant predictor of future negative returns with a coefficient of -0.071% and a t-statistic of 2.81, but retail net buying loses prediction with a t-value of only 0.99.

In Panel B, results are similar when I use last month turnover to divide the sample. In columns (1) and (2), for illiquid stocks, retail investors' trades have larger predictive powers for next month's returns than retail short selling does. However, the result is different for medium-liquid stocks in columns (3) and (4) and for liquid stocks in columns (5) and (6). Retail shorts have larger and more significant predictions on stock pricing than other retail investors' trades do. For example, in column (5) for the liquid stocks, *rtloib_total_w* retains predictive power at the 5% level, which is, however, fully explained by *rtlshort_w* rather than retail net buying *rtloib_w*, which can be seen in column (6). The standardized coefficient on *rtlshort_w* is 0.185% with a t-value of 4.87. In contrast, the coefficient on *rtloib_w* is only 0.044% with a t-value of 1.50 only. This suggests that non-shorting retail investors are more informed about illiquid stocks, while retail short sellers are more informed about liquid stocks. Short selling constraints and large fixed costs in gathering information could deter retail short sellers from acquiring the information on the illiquid stocks. Instead, retail short sellers actively trade the liquid stocks and profit from their mispricing.

To sum up, the empirical results in Table 5 suggest that retail short sales have predictive power beyond that coming from other retail traders as a group. Retail short sellers, however, can access different information that helps them generate greater abnormal returns.

2.5 Strategies of Retail Short Selling

2.5.1 Information-based Strategies

My earlier regression results suggest that retail shorting can predict stock returns. This raises the question why, and I examine a few alternative explanations. The first story is that retail short sellers are informed traders, similar to institutional short sellers as a group that other studies have looked at. Short sellers either have the ability to anticipate public news and better analyze publicly available information (Engelberg, Reed, and Ringgenberg, 2012) or can uncover new information themselves (e. g., Christophe, Ferri and Angel, 2004; Boehmer, Jones, and Zhang, 2015). Therefore, it is possible that retail short sellers' trading strategies exploit public information in this way. In this section, I identify and separate out the days when there is news announcement, and test whether retail short sellers' predictive power changes around these news days.

2.5.11 News Sample

I include three sets of news that are either earnings-related or analyst-related.

I select quarterly earnings announcements from Compustat. I compare the earnings announcements dates from I/B/E/S and Compustat and choose the earlier ones if they are different. My sample has 5,158 stocks with earnings announcements. Around 1.55% of days have such an announcement. Next, I define the unexpected earnings surprise, UE. UE is calculated as the announced EPS for a quarter less the corresponding consensus EPS forecast. Around 0.49% of days announce negative unexpected earnings surprises (UE<0).

The second news category consists of analyst recommendation changes from I/B/E/S. My sample has 3,570 stocks with analyst recommendation changes. Changes occur on average on 0.94% of sample days. I select negative news when an analyst recommendation change is downward. Negative analyst recommendation changes occur on average on 0.55% of sample days.

The third news category are analyst forecast revisions from I/B/E/S. My sample contains 4,528 stocks that have available analyst forecast revisions. An average 5.47% of days have revision announcements and these are negative, on average, on 3.03% of sample days.

Overall, 6.44% of all days are news days. On these news days, retail short sell volume accounts for 11.9% of total volume. 3.58% of all days are negative news days, and retail short sell volume is on average 7.37% of total volume on these negative news days. This suggests that a significant portion of retail short sales concentrates on news announcement dates.

2.5.12 Regression Setting

Next, I separate out the days with earning or analyst-related information to study short sellers' information. I introduce a dummy variable *news*, which equals one if day t=0 has an earning announcement, analyst recommendation change, or analyst forecast revision. I regress both next week's returns [t+1, t+5] and next month's returns [t+1, t+20] on weekly retail short sales, the news dummy, the interaction term of news with weekly retail short sells, and controls. The Fama-Macbeth regression is as follows:

$$r_{i,t+1,t+k} = b_0 + (b_1 + b_2 * News_t) * rtlshort_{i,t-5,t-1} + b_3 * News_t + r * x_{i,t-1} + e_{i,t} \quad (2)$$

Table 6 Panel A uses all the news types, either positive news or negative news. The dependent variable in the first three columns is next week's return. Column (1) shows the baseline results: the coefficient on weekly retail short sales is -0.018%, with a t-statistics of 3.01. Column (2) divides trading days into news days and non-news days. For overall retail shorting, the coefficient is -1.6 basis points, but on days with earning or analyst news, the coefficient is (-1.6)+(-1.8)=-3.4 basis points, which is more than double compared to the days without news. The incremental effect on earnings/analyst news days is also strongly statistically significant, with a t-statistic of 3.71. In column (3), I additionally introduce a *news* dummy to control for possible fixed time effects. The results are similar: the prediction on news days is (-1.6) +(-1.5) =-3.1 basis points, which almost doubles than the average effect of -1.6. But the predictions are quite short in the sense that the marginal difference between news days and non-news days becomes less significant in the longer term: when I regress next month's return on the same variables in column (4), (5), and (6), the difference between news days and non-news days becomes less significant.

There is another way to evaluate the importance of earnings' and analysts' news. One can use the fact that 6.44% of the days in the sample have an earnings-related or analyst-related announcement. That means, the overall underperformance for next week's return associated with a one-standard-deviation increase in retail short sales is:

$$6.44\% * (1.6+1.8) + (1 - 6.44\%) * 1.6 = 1.75 \text{ basis points per day}$$

The first term reflects the portion of short-sellers' information associated with earnings and analyst announcement days. In this case, it covers around 12.5% of overall underperformance.

Panel B uses only negative news. The incremental effect on earning/analyst news days is much larger in magnitude and in significance for both next week's return and next month's return. For example, in column (2), the coefficient for overall retail shorting is -1.4 basis points and only marginally significant; but on days with negative earning or analyst-related news, the coefficient is (-1.4)+(-6.0)=-7.4 basis points, which is more than five times larger compared to the days without negative news. The incremental effect is also statistically significant below the 1% level with a t value of 11.89. Even after additionally introducing a *news* dummy to control for the possible time fixed effects in column (3), the prediction on negative news days is (-1.6) +(-1.7) =-3.3 basis points, which more than doubles the average effect of -1.6. In column (5), when predicting next-month returns, the coefficient for overall retail shorting is -5.4 basis points. But on the days with negative news, the coefficient is (-5.4)+(-6.5)=-11.9 basis points, which more than doubles than the days without negative news. Even after introducing a dummy to control for time fixed effects in column (6), the difference between news days and non-news days is still -0.027% and statistically significant under 5% level.[8]

[8] I also select the days with positive news only. For retail short selling, the marginal difference between positive news days and non-positive news days in predicting future returns is not significantly different from zero.

Observing that 3.58% of days are related to negative information, I can calculate the percentage of underperformance related to negative information as:

$$0.014*(1-3.58\%)/[0.014*(1-3.58\%)+0.074*3.58\%]=16.3\%$$

Therefore, negative news announcements cover around 16.3% of all underperformance related to retail shorting. Overall, news events appear to have a strong effect on retail short sellers' ability to predict returns.

As a robustness test, I also test whether there exists abnormal retail shorting around news announcement dates. If retail shorting level significantly increases prior to news announcement dates, it suggests that they can actively predict information (Chirstophe, Ferri, and Angel, 2004; Christophe, Ferri, and Hsieh, 2010; Boehmer, Jones, and Zhang, 2015). If retail shorting level increases around news announcement dates (Engelberg, Reed, and Ringgenberg, 2010), it suggests that they can better analyze information. Actually, in unreported results, I find that retail shorting only increases on and after news announcement dates, consistent with the point that retail short sellers can better process information.

2.5.13 *Regression Setting Comparing Retail Short Selling vs. Other Retail Trading*

In this section, I compare the information advantage of retail short selling to that of other retail trading using cross-sectional regressions. I ask whether it is possible that non-shorting retail traders can exploit public information as well.

Similar to the setting above, I identify the days when there are earnings- or analyst-related information events on day t=0. In column (1), I regress next week's return [t+1, t+5] on total retail order imbalance *rtloib_total_w* [t-5, t-1], the news dummy (t=0), the interaction term of news with weekly retail order imbalances, and controls. In column (2), I further divide total retail order imbalance into retail short selling *rtlshort_w* [t-5, t-1] and retail net buying *rtloib_w* [t-5, t-1], and then regress next week's return on retail short selling, retail net buying, the news dummy, the interaction term of news with retail shorting, the interaction term of news with retail net buying, and controls. By comparing the interacted coefficients, we can evaluate the relative incremental stock returns associated with previous week's retail trading/shorting activity that is due to earnings news or analyst changes.

Table 7 reports results. In column (1), the coefficient on the interaction term is not significantly different from zero, suggesting that on average, non-short retail investors cannot anticipate and exploit information. In column (2), the coefficient on the interaction term of retail shorting with news dummy is -0.018% with a t-statistic of 2.58, significant under 1% level. In contrast, the coefficient on the interaction term of retail net buying with news dummy is statistically indistinguishable from zero with a t-statistic of 0.74 only. The results are consistent with retail short selling conveying novel information. In contrast, there is no evidence that other retail traders act on such information.

2.5.2. Liquidity Provision Strategies

Retail traders can also use short sales to provide liquidity to the impatient who require immediacy. As the buying pressure subsides, prices will revert to the fundamental values and short sellers can cover their positions and potentially earn positive returns. In this scenario, the trading patterns and return predictability are the result of short sellers receiving compensation for providing immediacy (Nagel, 2012).

The literature is divided on this point. On one hand, recent studies by Kaniel, Saar, and Titman (2008), Kaniel, Liu, Saar, and Titman (2012), and Barrot, Kaniel, and Sraer (2016) find that retail investors as contrarian individuals can provide liquidity when market liquidity dries up. But none of these studies differentiates retail shorts from other types of retail trading. Retail short selling is different from the typical retail trading in that short sellers will incur costs from foregone interest on collateral and stand risks from recalling shares lent out. On the other hand, Diether, Lee, and Werner (2009) and Comerton-Forde, Jones, and Putnins (2016) verify that short sellers can actually provide liquidity to the impatient buyers. But most of these samples are about institutional short sells. So, in this section, I fill this gap and test whether retail short sellers also provide liquidity.

2.5.11 Are Retail Short Sellers Contrarian or Momentum?

I start by analyzing how retail short sales react to past returns. The dependent variable is the share of retail short sales on day t=0 *rtlshort*. I use Fama Macbeth regressions and various horizons of returns, including contemporaneous same-day returns *ret(0)*, past week returns [t-5, t-1] *ret(w-1),* and past month returns [t-26, t-6] *ret(m-1).* For other control variables, I use log market cap, log book-to-market ratio, turnover, and daily return volatility, all observed from the previous month-end. To control for auto-correlation in retail short sells, I also include its own lag, weekly retail short sells [t-5, t-1], or *rtlshort_w*. The last independent variable is bi-monthly short interest from Compustat *shortint*, which can proxy for the overall shorting on stock exchanges. I select the most recent short interest that can be observed on day t=0. That means the most recent short interest is reported on day t-1. To facilitate comparisons, I standardize both the dependent variable and the independent variables. The regression model is as:

$$rtlshort_{i,t=0} = b_0 + b_1 * ret_{i,0} + b_2 * ret_{i,t-5,t-1} + b_3 * ret_{i,t-26,t-6} + r * x_{i,t-1} + e_{i,t} \quad (3)$$

Table 8 presents results. In column (1), independent variables only include returns. The coefficients on contemporaneous return, past week returns, and past month returns (b_1, b_2, b_3) are 0.014, 0.019 and 0.011, respectively. All three coefficients are positive and statistically significant, suggesting that retail short sellers are strong contrarians, who sell the stocks after prices rise. In column (2), I additionally include firm level characteristics from the previous month-end in the regression. The coefficients on the returns continue to be significant. Similar results are presented in column (3), when I control for past retail short sales and short interest as well. On average, more retail shorting tends to follow small stocks, growth stocks, more liquid stocks, and more volatile stocks.

2.5.12 How Do Retail Short Sells React to Changes in Buying Pressure?

It is possible that past and contemporaneous price increases are caused by factors that themselves trigger short selling activity. For example, greater past returns can arise when buying pressure increases during this period. But this effect would be purely liquidity-motivated rather than information-based, if retail short sellers just step in to provide liquidity to the impatient buy orders. To better understand this relation, I use three variables that help control for buying pressure at the stock level.

The first measure is based on Lee and Ready (1991). I classify trades as either buyer-initiated or seller-initiated, and then calculate order imbalance as: *LRoib*= (*vol_buy-vol_sell*)/*vol*. I construct an additional measure from these imbalances that differentiates between positive and negative imbalances. Following Diether, Lee, and Werner (2009) and Boehmer, Jones, and Zhang (2008), I define this measure as *LRoib+*, which equals *LRoib* if *LRoib*>0 otherwise zero. *LRoib+_w* includes positive buy-order imbalances in previous week.

The second measure is retail net buying. Similar to above, I calculate the retail order imbalance as (*rtlvol_buy*-(*rtlvol_sell-rtlshort_vol*))/*vol*, and *rtloib+* equals *rtloib* if *rtloib*>0 and zero otherwise. *rtloib+_w* includes positive retail buy net buying over the previous week.

The third measure is the imbalance based on "other" trades (other than retail), or *otheroib+*. To construct this measure, I subtract retail trades from overall trades, so this measure also captures institutional trades. *Otheroib* is calculated as ((*vol_buy-rtlvol_buy*)-(*vol_sell-rtlvol_sell*))/*vol*, and *otheroib+* equals *otheroib* if *otheroib*>0 and zero otherwise. *Otheroib+_w* includes positive other buy-order imbalances over the previous week.

Before partitioning into positive *oib+* and negative *oib-*, I standardize order imbalance so that both the dependent and the other independent variables to have zero mean and unit standard deviation each day. Table 9 presents cross-sectional regression results of retail short sells on contemporaneous and past buying pressures, controlling for the returns, log market cap, log book-to-market ratio, turnover, volatility, last week retail shorting, and the most recent short interest.

Column (1) shows that today's retail short selling is positively correlated with both contemporaneous and past buy order imbalances, which is consistent with a liquidity provision story. The coefficients on *LRoib+* and *LRoib+_w* are 0.014 and 0.010, respectively. Both are significant at the 1% level. In column (2), retail short sells are also positively related to contemporaneous and past retail buy order imbalances. The coefficients on *rtloib+* and *rtloib+_w* are 0.021 and 0.032 with t-statistics of 7.31 and 34.90. In column (3), where I introduce other buy order imbalances which are most likely to reflect institutional buying pressures, retail short sells are significantly positively correlated with contemporaneous "other" buy-order imbalances in a much larger magnitude. The coefficient on *otheroib+* is 0.153 with a t-statistic of 27.84. There is no evidence, however, that retail short sales are related to other order imbalances in the past. In column (4), when I put retail order imbalances and "other" order imbalances together in the same regression, retail short sells have a much larger relation with institutional order imbalances (magnitude of 0.149 and t value of 27.57) than with retail order imbalances (magnitude of 0.025 and t value of 9.16).

In sum, the estimates in Table 9 show that with greater (institutional) buying pressure, retail short sellers will increase their positions and step in to provide liquidity. Especially, in unreported results, institutional buying pressures are related to the subsequent negative returns. Combining this with the prior results that greater retail short selling is associated with future negative stock returns in Table 2 and Table 4 as well, my results suggest that retail short sellers receive compensation for providing liquidity to institutions that need to execute their trades immediately, consistent with Kaniel et al. (2008).

2.5.3. Discussion

So far, I have presented evidence that retail short sellers are informed traders but can switch to become liquidity providers when needed by aggressive buyers. This prompts the question when retail traders' embrace information trading and when they switch to liquidity providers.

To address this question, I decompose the relation of retail short selling activities to past returns into days with news announcement and days without news, when return reversals are known to be larger (Tetlock, 2010). I interact the *news* dummy on day t=0 with contemporaneous and past returns. The interaction terms could reflect the difference between liquidity provisions' trading on news and non-news days. The regression is as follows:

$$rtlshort_{i,t=0} = b_0 + b_1 * ret_{i,t=0} * news_{i,t=0} + b_2 * ret_{i,t-5,t-1} * news_{i,t=0} + b_3 * ret_{i,t-26,t-6} * news_{i,t=0} + b_4 * ret_{i,t=0} + b_5 * ret_{i,t-5,t-1} + b_6 * ret_{i,t-26,t-6} + r * x_{i,t-1} + e_{i,t} \quad (4)$$

Table 10 Panel A reports the results. In column (1), I use all news types. Without news, the relation of retail shorting to contemporaneous return (b_4) is 0.016. The positive correlation suggests that retail short sellers are strong contrarians. But on the days with news announcement, this relation decreases to 0.010 ($b_1 + b_4$). The coefficient on the interaction term of news dummy with contemporaneous return (b_1) is -0.006, implying that on news release days, retail short sellers become less contrarian relative to returns. The relation between retail shorting and past returns also decreases significantly on news release days. Similar results apply to Column (2) where I only select the days of negative news, including negative unexpected earning surprise, or analyst recommendation downgrades, or analyst negative revision of forecasts. On negative news release days, the relation of retail shorting with contemporaneous return decreases to 0.08 (0.017-0.009), suggesting that retail short sellers are less contrarian relative to past returns when they are accompanied by negative news. However, results are different in column (3). When I only select positive news announcement days, all the interaction terms (b_1, b_2, b_3) are insignificant from zero, suggesting that retail short selling can't exploit positive information, so that it won't affect retail traders liquidity provision strategies.

I also separate days when the market is more one-sided with greater temporary buying pressure. As above, I continue to use positive order imbalances to proxy for buying pressures on individual stocks. I interact positive order imbalances with contemporaneous and past returns, which can reflect the difference between liquidity provisions' trading on stocks with or without liquidity shocks. The regression is as follows:

$$rtlshort_{i,t=0} = b_0 + b_1 * ret_{i,t=0} * oib(+)_{i,t=0} + b_2 * ret_{i,t-5,t-1} * oib(+)_{i,t=0} + b_3 * ret_{i,t-26,t-6} * oib(+)_{i,t=0} + b_4 * ret_{i,t=0} + b_5 * ret_{i,t-5,t-1} + b_6 * ret_{i,t-26,t-6} + r * x_{i,t-1} + e_{i,t} \quad (5)$$

Table 10 Panel B reports results. In column (1), I use *LRoib+* to proxy for overall buying pressure, which equals *LRoib* if *LRoib*>0 otherwise zero. All the interaction terms (b_1, b_2, b_3) are positively significant. The average relation of retail shorting with contemporaneous return (b_4) is 0.004, but for the stocks with buying pressure, the relation increases to 0.018 (0.004+0.014). This suggests that retail short sellers become more contrarian relative to the returns on stocks with greater temporary buying pressures. Column (2) uses positive retail net buying to proxy for retail buying pressure and column (3) uses positive "other" net buying to proxy for institutional buying pressure. The interaction terms (b_1) are both positively significant, but larger for institutional buying pressure than for retail buying pressures. In column (3), the average relation of retail shorting with contemporaneous return (b_4) even becomes negative of -0.003, suggesting that retail short sales are not contrarian in the short term for the stocks without impatient institutional buyers. But for the stocks with larger institutional buying pressure, the relation increases to 0.007 (0.010-0.003).

Overall, the results in Table 10 suggest that retail short sellers tend to use different trading strategies. They provide liquidity, possibly in a strategic manner, and act as contrarian traders, especially when there is a temporary (institutional) buying pressure. But at other times, retail short sellers have private information. In these cases, retail traders behave more aggressively and are less contrarian relative to past returns.

2.6. Conclusions

In this paper, I provide the most extensive evidence thus far about the return predictability of retail short selling. Exploiting the short sell transaction data provided by FINRA, I select a large group of short sell transactions initiated by retail investors between January 2010 and December 2016.

I characterize the trading patterns and strategies of retail short sellers and find that retail short selling is a strong predictor of negative stock returns. A portfolio that mimics weekly retail shorting earns a risk-adjusted value-(equal-) weighted return of 0.024% (0.049%) in the next 20 trading days, annualized at 6% (12.25%). Using a formal cross-sectional regression, I find that most of the predictive power survives after controlling for firm characteristics and the activities of other informed traders.

My study contributes to the literature about the informativeness of retail investors. Recent literature finds that there is a positive relation between retail investors' trading and future returns (Kaniel, Saar, and Titman 2008; Kelley and Tetlock 2013; Boehmer, Jones, and

Zhang, 2017). I contribute to this literature. Among other things, in this paper, I can differentiate *smart* transactions by retail short sellers from total retail trading. I find that the trading patterns of these retail short sellers have distinct predictive powers that go beyond the information that retail traders have as a group. Retail short sellers tend to use different trading strategies depending on market conditions. When there is greater buying pressure and the market is rather one-sided, retail short sellers step in to provide liquidity and earn compensation for it. At other times, retail short sellers act more aggressively to exploit their information. In contrast, non-short retail orders cannot act on such information. This, retail short sales are integral to both price discovery and liquidity provision.

My study also contributes to the literature about the different types of short sellers and their strategies. Boehmer, Jones, and Zhang (2008) find that about 75% of all short sales are executed by institutions while individuals only represent less than 2%. As a group, retail investors may face high lending fees and entrance restrictions when short selling. Therefore, only the more sophisticated retail investors will initiate short selling activities. My paper is consistent with this view. I provide empirical results showing that retail short sellers can indeed exploit public information and predict returns. But this is not all—retail short sellers also have their own private information that is distinct from the information that other retail investors have in general.

Overall, my paper strengthens recently emerging evidence that retail traders are quite informed traders. But their contribution is not limited to supporting price discovery; they also fulfil important liquidity provision roles in U.S. equity markets.

References

Barber, B.M. and Odean, T., 2000. Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance*, *55*(2), pp.773-806.

Barber, B.M. and Odean, T., 2001. Boys will be boys: Gender, overconfidence, and common stock investment. *The quarterly journal of economics*, *116*(1), pp.261-292.

Barber, B.M. and Odean, T., 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies*, *21*(2), pp.785-818.

Barrot, J.N., Kaniel, R. and Sraer, D., 2016. Are retail traders compensated for providing liquidity? *Journal of Financial Economics*, *120*(1), pp.146-168.

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

Boehmer, E., Jones, C. and Zhang, X., 2015. What do short sellers know?

Boehmer, E., Jones, C. and Zhang, X., 2017. Tracking retail investor activity.

Chordia, T. and Subrahmanyam, A., 2004. Order imbalance and individual stock returns: Theory and evidence. *Journal of Financial Economics*, *72*(3), pp.485-518.

Christophe, S.E., Ferri, M.G. and Angel, J.J., 2004. Short - selling prior to earnings announcements. *Journal of Finance*, *59*(4), pp.1845-1876.

Christophe, S.E., Ferri, M.G. and Hsieh, J., 2010. Informed trading before analyst downgrades: Evidence from short sellers. Journal of Financial Economics, 95(1), pp.85-106.

Comerton-Forde, C., Jones, C.M. and Putniņš, T.J., 2016. Shorting at close range: a tale of two types. *Journal of Financial Economics*, *121*(3), pp.546-568.

Comerton-Forde, C., Malinova, K. and Park, A., 2018. Regulating dark trading: Order flow segmentation and market quality. *Journal of Financial Economics*, *130*(2), pp.347-366.

Diether, K.B., Lee, K.H. and Werner, I.M., 2008. Short-sale strategies and return predictability. *Review of Financial Studies*, *22*(2), pp.575-607.

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

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

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

Gamble, K.J. and Xu, W., 2017. Informed retail investors: Evidence from retail short sales. *Journal of Empirical Finance*, *40*, pp.59-72.

Jain, A. and Jain, C., 2017. Is Short Selling in OTC Stocks Informative?

Jegadeesh, N. and Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, *48*(1), pp.65-91.

Kaniel, R., Liu, S., Saar, G. and Titman, S., 2012. Individual investor trading and return patterns around earnings announcements. *Journal of Finance*, *67*(2), pp.639-680.

Kaniel, R., Saar, G. and Titman, S., 2008. Individual investor trading and stock returns. *Journal of Finance*, *63*(1), pp.273-310.

Kelley, E.K. and Tetlock, P.C., 2013. How wise are crowds? Insights from retail orders and stock returns. *Journal of Finance*, *68*(3), pp.1229-1265.

Kelley, E.K. and Tetlock, P.C., 2017. Retail short selling and stock prices. *Review of Financial Studies*, *30*(3), pp.801-834.

Kwan, A., Masulis, R. and McInish, T.H., 2015. Trading rules, competition for order flow and market fragmentation. *Journal of Financial Economics*, *115*(2), pp.330-348.

Lee, C. and Ready, M.J., 1991. Inferring trade direction from intraday data. *Journal of Finance*, *46*(2), pp.733-746.

Nagel, S., 2012. Evaporating liquidity. *Review of Financial Studies*, *25*(7), pp.2005-2039.

Odean, T., 1998. Are investors reluctant to realize their losses? *Journal of Finance*, 53(5), pp.1775-1798.

Reed, A.V., Samadi, M. and Sokobin, J.S., 2017. Where do shorts trade? Intermarket competition and information horizon.

Tetlock, P.C., 2010. Does public financial news resolve asymmetric information? *Review of Financial Studies*, *23*(9), pp.3520-3557.

Zhu, H., 2014. Do dark pools harm price discovery?. *The Review of Financial Studies*, *27*(3), pp.747-789.

Table 1: Summary Statistics

This table presents summary statistics of U.S. common stocks listed on NYSE, AMEX and Nasdaq from January 2010 to December 2016 with a share price of at least $1. In Panel A, across all stocks and all days, I first report the pooled sample mean for traded shares (*vol*), number of trades (*trd*), buyer-initiated traded shares (*vol_buy*), seller-initiated traded shares (*vol_sell*), number of buyer-initiated trades (*trd_buy*), and number of seller-initiated trades (*trd_sell*). Data is from daily TAQ. Buys and sells are identified following Lee and Ready (1991) algorithm. Panel B reports off-exchange trading and retail trading. It includes shares traded off-exchanges (*otcvol*), number of trades off-exchanges (*otctrd*), retail buy volumes (*rtlvol_buy*), retail sell volumes (*rtlvol_sell*), retail buy trades (*rtltrd_buy*), and retail sell trades (*rtltrd_sell*). OTC trading is from daily TAQ with exchange code "D". Retail buys and sells are identified as the OTC trading with sub-penny price improvement. Panel C reports OTC short selling volumes (*otcshort_vol*) and OTC short selling trades (*otcshort_trd*). OTC short selling data is from FINRA. Panel C also reports *retail short selling* volumes (*rtlshort_vol*) and *retail short selling* trades (*rtlshort_trd*). Retail short selling is identified as the group of OTC short selling which has sub-penny price improvements compared with round penny. In Panel D**,** I calculate the percentage of retail shorting volumes/trades in total retail trading volumes /trades (*rtlshort_rtlvol* /*rtlshort_rtltrd*), the percentage of retail shorting volumes/ trades in total OTC shorting volumes/trades (*rtlshort_otcshortvol /rtlshort_otcshorttrd*), the percentage of retail shorting volumes in all the shares traded (*rtlshort*), and the percentage of retail shorting trades in all the number of trades (*rtlshort_t*).

Variable	N	Mean	Std Dev	Median	Q1	Q3
Panel A						
Vol	6224866	1,148,048	6,314,922	188,166	37,456	741,700
Trd	6224866	5,603	13,684	1,295	220	5,049
Vol_buy	6224866	569,372	3,000,150	91,177	17,297	366,708
Vol_sell	6224866	578,089	3,349,302	94,383	18,655	372,891
Trd_buy	6224866	2,802	6,879	639	105	2,518
Trd_sell	6224866	2,801	6,842	649	110	2,525
Panel B						
Otcvol	6224866	393,715	2,626,586	56,599	11,295	230,794
Otctrd	6224866	1,296	3,660	284	54	1,083
Rtlvol_buy	6224866	42,526	275,679	4,758	908	19,915
Rtlvol_sell	6224866	42,177	258,384	5,140	1,028	20,871
Rtltrd_buy	6224866	111	420	21	4	78
Rtltrd_sell	6224866	107	362	23	5	81
Panel C						
Otcshort_vol	6224866	160,189	1,010,098	19,049	3,233	88,155
Otcshort_trd	6224866	583	1,657	107	16	472
Rtlshort_vol	6224866	7,925	54,519	800	0	3,948
Rtlshort_trd	6224866	24	85	4	0	18
Panel D						
Rtlshort_rtlvol	5889923	0.1092	0.1380	0.0704	0.0075	0.1520
Rtlshort_rtltrd	5889923	0.1138	0.1283	0.0833	0.0169	0.1622
Rtlshort_otcshortvol	5840798	0.0636	0.1099	0.0333	0.0053	0.0736
Rtlshort_otcshorttrd	5840798	0.0545	0.0929	0.0319	0.0077	0.0636
Rtlshort	6224866	0.0078	0.0221	0.0032	0.0000	0.0079
Rtlshort_t	6224866	0.0053	0.0146	0.0025	0.0000	0.0056

Table 2: Portfolio Returns Based on Retail Short Selling

This table presents daily returns for portfolios based on weekly retail short sells from January 2010 to December 2016. Each day, stocks are sorted into quintiles based on weekly retail short selling volumes *(rtlshort_vol_w)* in Panel A, weekly retail short selling trades (*rtlshort_trd_w*) in Panel B, weekly retail short selling's percentage of volumes (*rtlshort_w*) in Panel C, and weekly retail short selling's percentage of trades (*rtlshort_t_w*) in Panel D. After skipping 1 day, value-(equal-) weighted portfolios are held for 20 trading days. This process is repeated each trading day, so each trading day's portfolio return is an average of 20 different portfolios, with 1/20 of the portfolio rebalanced each day, following Jegadeesh and Titman (1993). Daily calendar time returns and Fama and French (1993) three factor alphas are reported in percent. Newey-west (1987) t-statistics based on five lags appear in parentheses.

Panel A: Number of retail short sell volume (*rtlshort_vol_w*)				
	Value-weighted		Equal-weighted	
	Return	Alpha	Return	Alpha
1 (Least)	0.058	0.022	0.076	0.053
2	0.061	0.016	0.061	0.020
3	0.059	0.010	0.055	0.006
4	0.056	0.005	0.046	-0.009
5(Most)	0.048	-0.002	0.035	-0.023
5-1	-0.010	-0.024***	-0.041**	-0.075***
	(-0.998)	(-4.505)	(-2.528)	(-7.523)

Panel B: Number of retail short sell trades (*rtlshort_trd_w*)				
	Value-weighted		Equal-weighted	
	Return	Alpha	Return	Alpha
1 (Least)	0.059	0.026	0.078	0.057
2	0.062	0.018	0.059	0.019
3	0.059	0.011	0.053	0.002
4	0.056	0.005	0.045	-0.010
5(Most)	0.048	-0.002	0.039	-0.018
5-1	-0.010	-0.027***	-0.038**	-0.074***
	(-0.997)	(-4.985)	(-2.415)	(-7.295)

Panel C: Retail shorting's percentage of volume (*rtlshort_w*)				
	Value-weighted		Equal-weighted	
	Return	Alpha	Return	Alpha
1 (Least)	0.061	0.017	0.073	0.045
2	0.059	0.010	0.058	0.009
3	0.050	0.002	0.053	0.002
4	0.045	-0.006	0.047	-0.005
5(Most)	0.049	-0.006	0.041	-0.004
5-1	-0.012	-0.024***	-0.032***	-0.049***
	(-1.185)	(-2.606)	(-2.819)	(-6.865)

Panel D: Retail shorting's percentage of trades (*rtlshort_t_w)*				
	Value-weighted		Equal-weighted	
	Return	Alpha	Return	Alpha
1 (Least)	0.061	0.015	0.073	0.043
2	0.056	0.007	0.060	0.010
3	0.050	0.001	0.052	0.000
4	0.046	-0.006	0.046	-0.005
5(Most)	0.043	-0.011	0.041	-0.002
5-1	-0.018	-0.027**	-0.032***	-0.045***
	(-1.556)	(-2.378)	(-3.052)	(-6.294)

Table 3: Portfolio Returns Based on Retail Trading and Retail Short Selling

This table presents daily value-weighted returns for portfolios first sorted on weekly retail order imbalances, and then second sorted on weekly retail short sells. Sample period is from January 2010 to December 2016. Each day, stocks are first sorted into quintiles based on the weekly retail order imbalances, calculated as: (retail buy volumes-retail *long sell* volumes)/total volumes, over the previous five trading day. Within each quintile, stocks are second sorted into quintiles based on retail short sell percentage of volumes over the previous five trading day. After skipping 1 day, value-weighted portfolios are held for 20 trading days. I adjust overlaps following Jegadeesh and Titman (1993). Daily calendar time returns and Fama and French (1993) three factor alphas are reported in percent. Newey-west (1987) t-statistics based on five lags appear in parentheses.

Second sort: Weekly retail short sell percentage of volumes (*rtlshort_w*)

	First sort: Weekly retail order imbalances (*rtloib_w*)									
	Value-weighted Return					Value-weighted Alpha				
	1 (Low)	2	3	4	5 (High)	1 (Low)	2	3	4	5 (High)
1 (Least)	0.045	0.058	0.064	0.069	0.078	0.014	0.015	0.015	0.021	0.044
	(2.462)	(2.870)	(2.855)	(3.190)	(3.958)	(2.319)	(3.021)	(3.189)	(4.445)	(5.923)
2	0.047	0.057	0.063	0.058	0.066	0.006	0.009	0.012	0.010	0.021
	(2.354)	(2.641)	(2.936)	(2.857)	(3.194)	(1.371)	(2.149)	(3.479)	(2.798)	(3.150)
3	0.050	0.051	0.052	0.047	0.055	0.004	0.003	0.002	-0.004	0.007
	(2.371)	(2.470)	(2.543)	(2.212)	(2.660)	(0.797)	(0.848)	(0.668)	(-1.244)	(1.034)
4	0.035	0.044	0.048	0.044	0.041	-0.011	-0.003	-0.003	-0.010	-0.015
	(1.641)	(2.236)	(2.285)	(1.949)	(1.606)	(-2.041)	(-0.802)	(-0.923)	(-2.176)	(-1.438)
5 (Most)	0.033	0.041	0.042	0.041	0.054	-0.010	-0.007	-0.010	-0.015	-0.005
	(1.397)	(1.914)	(1.929)	(1.630)	(1.711)	(-1.594)	(-1.502)	(-2.548)	(-1.884)	(-0.273)
5-1	-0.012	-0.017***	-0.022***	-0.027**	-0.024	-0.025***	-0.022***	-0.025***	-0.036***	-0.050**
	(-1.266)	(-2.695)	(-3.529)	(-2.488)	(-1.025)	(-3.358)	(-3.330)	(-3.838)	(-3.397)	(-2.384)

Table 4: Cross-sectional Regressions of Returns on Retail Short Selling

This table presents daily Fama Macbeth (1973) regressions of next month's returns [t+1, t+20] on weekly retail short selling [t-5, t-1] (*rtlshort_w)* and control variables measured as of day t. Sample period is from January 2010 to December 2016. Dependent variable in Panel A is cumulative raw returns [t+1, t+20], and dependent variable in Panel B is cumulative risk-adjusted returns [t+1, t+20] using Fama-French three factor model, with betas estimated from previous year. I also include weekly dark pool non-retail short sells [t-5, t-1] (*darkshort_w),* calculated as (OTC short sell volumes-retail short sell volumes)/total volumes, and weekly retail order imbalance [t-5, t-1] (*rtloib_w),* calculated as (retail buy volumes-retail *long sell* volumes)/total volumes. Other control variables include previous week return [t-5, t-1] *ret(w-1),* previous month returns [t-25, t-6] *ret (m-1),* and previous year returns [t-251, t-26] *ret(y-1).* I also include log market cap (*size*), log book to market ratio (*btm*), monthly turnover (*turn*), and monthly volatility of daily returns (*vol*), all measured at the end of previous month. To account for serial correlation in the coefficients, the standard deviation of time-series is adjusted using Newey-west with 20 lags. Dependent variables are in percent, and all the independent variables are standardized each day t.

Panel A:

	Y= Cumulative raw returns [t+1, t+20]			
	1	2	3	4
Rtlshort_w	-0.114***	-0.065***	-0.051***	-0.073***
	(-4.04)	(-3.06)	(-2.58)	(-3.50)
Darkshort_w			-0.085***	
			(-2.74)	
Rtloib_w				0.126***
				(11.70)
Log(Size)		0.109	0.106	0.107
		(1.42)	(1.38)	(1.39)
Log(Btm)		0.113	0.106	0.114
		(1.54)	(1.45)	(1.56)
Turn		-0.170***	-0.158***	-0.172***
		(-3.44)	(-3.35)	(-3.48)
Vol		-0.343***	-0.337***	-0.342***
		(-4.82)	(-4.80)	(-4.80)
Ret(w-1)		-0.143***	-0.137***	-0.148***
		(-3.98)	(-3.74)	(-4.13)
Ret (m-1)		0.048	0.046	0.051
		(0.82)	(0.78)	(0.87)
Ret (y-1)		0.157**	0.152**	0.156**
		(2.07)	(2.01)	(2.06)
Intercept	1.013**	1.013**	1.013**	1.013**
	(2.57)	(2.57)	(2.57)	(2.57)
R square	0.001	0.041	0.042	0.042
# of days	1736	1736	1736	1736
Obs	5542587	5542587	5542587	5542587

Panel B:

	Y= Cumulative risk-adjusted returns [t+1, t+20]			
	1	2	3	4
Rtlshort_w	-0.070***	-0.056***	-0.038**	-0.065***
	(-2.93)	(-3.26)	(-2.41)	(-3.81)
Darkshort_w			-0.122***	
			(-3.97)	
Rtloib_w				0.118***
				(10.74)
Log(Size)		-0.223***	-0.224***	-0.224***
		(-3.32)	(-3.31)	(-3.34)
Log(Btm)		0.127***	0.117***	0.129***
		(3.12)	(2.84)	(3.15)
Turn		-0.197***	-0.183***	-0.199***
		(-4.26)	(-4.06)	(-4.30)
Vol		-0.294***	-0.287***	-0.293***
		(-5.10)	(-5.04)	(-5.08)
Ret(w-1)		-0.141***	-0.132***	-0.146***
		(-4.30)	(-3.95)	(-4.46)
Ret (m-1)		0.042	0.040	0.045
		(0.78)	(0.74)	(0.84)
Ret (y-1)		0.161**	0.154**	0.160**
		(2.57)	(2.47)	(2.56)
Intercept	0.119**	0.117**	0.117**	0.117**
	(2.04)	(2.02)	(2.02)	(2.02)
R square	0.001	0.029	0.030	0.029
# of days	1736	1736	1736	1736
Obs	5364476	5364476	5364476	5364476

Table 5: Sub-sample Cross-sectional Regressions

This table presents sub-sample daily Fama Macbeth (1973) regressions by firm market capitalization in Panel A and by firm liquidity in Panel B. Sample period is from January 2010 to December 2016. I first sort all the stocks into 3 groups based on previous month-end firm size/firm turnover, and then estimate Fama Macbeth (1973) regressions for each subgroup. Dependent variable is cumulative raw returns [t+1, t+20], and independent variables include total weekly retail order imbalance [t-5, t-1] (*rtloib_total_w*), which is (retail buy volumes-retail *sell* volumes)/total volumes, weekly retail short sells [t-5, t-1] (*rtlshort_w*), and weekly retail order imbalance [t-5, t-1] (*rtloib_w),* which is (retail buy volumes-retail *long sell* volumes)/total volumes. Other control variables include previous week return [t-5, t-1] *ret(w-1)*, previous month returns [t-25, t-6] *ret (m-1),* and previous year returns [t-251, t-26] *ret(y-1).* They also include log market cap (*size*), log book to market ratio (*btm*), monthly turnover (*turn*), and monthly volatility of daily returns (*vol*), all measured at the end of previous month. To account for serial correlation in the coefficients, the standard deviations of time-series are adjusted using Newey-west with 20 lags. Dependent variables are in percent, and all the independent variables are standardized each day t.

Panel A: Market-cap groups

	Y= Cumulative returns [t+1, t+20]					
	Small		Medium		Large	
	1	2	3	4	5	6
Rtloib_total_w	0.189***		0.117***		0.023	
	(10.18)		(5.94)		(1.44)	
Rtlshort_w		-0.065***		-0.103***		-0.071***
		(-2.71)		(-4.27)		(-2.81)
Rtloib_w		0.182***		0.108***		0.017
		(9.81)		(5.28)		(0.99)
Log(Size)	0.229**	0.226**	0.050	0.042	-0.069	-0.062
	(2.08)	(2.06)	(1.04)	(0.88)	(-1.39)	(-1.23)
Log(Btm)	0.292***	0.292***	0.056	0.053	0.001	-0.007
	(3.88)	(3.89)	(0.61)	(0.57)	(0.02)	(-0.11)
Turn	-0.279***	-0.278***	-0.194***	-0.190***	-0.171***	-0.159**
	(-4.34)	(-4.32)	(-2.73)	(-2.69)	(-2.69)	(-2.57)
Vol	-0.531***	-0.533***	-0.149*	-0.147*	0.010	0.013
	(-6.05)	(-6.08)	(-1.74)	(-1.72)	(0.12)	(0.15)
Ret(w-1)	-0.232***	-0.231***	-0.104**	-0.101**	-0.038	-0.035
	(-5.35)	(-5.32)	(-2.56)	(-2.49)	(-0.96)	(-0.90)
Ret (m-1)	0.112*	0.113*	-0.002	0.001	0.028	0.030
	(1.68)	(1.69)	(-0.03)	(0.01)	(0.41)	(0.44)
Ret (y-1)	0.228***	0.229***	0.105	0.109	0.071	0.074
	(2.60)	(2.60)	(1.22)	(1.27)	(0.86)	(0.90)
Intercept	0.710*	0.710*	1.154**	1.154**	1.176***	1.176***
	(1.85)	(1.85)	(2.51)	(2.51)	(3.13)	(3.13)
R square	0.045	0.046	0.052	0.054	0.081	0.083
# of days	1736	1736	1736	1736	1736	1736
Obs	1846963	1846963	1848103	1848103	1847521	1847521

Panel B: Turnover groups

	Y= Cumulative returns [t+1, t+20]					
	Low		Medium		High	
	1	2	3	4	5	6
Rtloib_total_w	0.184***		0.122***		0.065**	
	(10.75)		(5.03)		(2.33)	
Rtlshort_w		-0.046**		-0.138***		-0.185***
		(-2.05)		(-4.71)		(-4.87)
Rtloib_w		0.181***		0.107***		0.044
		(10.82)		(4.26)		(1.50)
Log(Size)	-0.009	-0.009	-0.022	-0.041	0.358***	0.325***
	(-0.09)	(-0.10)	(-0.29)	(-0.55)	(4.06)	(3.77)
Log(Btm)	0.222***	0.222***	0.091	0.087	-0.029	-0.043
	(3.67)	(3.68)	(1.38)	(1.33)	(-0.27)	(-0.41)
Turn	0.059	0.060	-0.016	-0.018	-0.292***	-0.278***
	(0.94)	(0.96)	(-0.57)	(-0.62)	(-4.76)	(-4.64)
Vol	-0.445***	-0.447***	-0.322***	-0.313***	-0.215**	-0.194**
	(-6.98)	(-6.99)	(-4.52)	(-4.41)	(-2.38)	(-2.19)
Ret(w-1)	-0.153***	-0.153***	-0.125***	-0.122***	-0.152***	-0.145***
	(-4.17)	(-4.17)	(-3.39)	(-3.30)	(-3.27)	(-3.13)
Ret (m-1)	0.040	0.040	0.082	0.085	0.014	0.014
	(0.74)	(0.72)	(1.37)	(1.42)	(0.18)	(0.19)
Ret (y-1)	0.215***	0.215***	0.174**	0.177**	0.106	0.122
	(3.48)	(3.48)	(2.38)	(2.42)	(1.05)	(1.19)
Intercept	1.120***	1.120***	1.148***	1.148***	0.773	0.773
	(3.36)	(3.36)	(2.89)	(2.89)	(1.63)	(1.63)
R square	0.044	0.045	0.051	0.053	0.059	0.061
# of days	1736	1736	1736	1736	1736	1736
Obs	1846963	1846963	1848103	1848103	1847521	1847521

Table 6: Cross-sectional Regression of Returns on Retail Shorting: News vs. Non-News

This table presents daily Fama Macbeth (1973) regressions of returns on retail short selling, by dividing days into news days and non-news days. Sample period is from January 2010 to December 2016. Panel A includes all the news, and Panel B includes negative news only. The news dummy equals to 1 if there is earning announcement, or analyst recommendation change, or analyst forecast revision on that day t=0, else equals to zero. The negative news dummy equals to 1 if there is negative unexpected earnings surprise, or analyst recommendation downgrades, or negative revisions of analyst forecasts on that day t=0, else equals to zero. The dependent variables are cumulative raw returns over [t+1, t+5] and over [t+1, t+20]. Independent variables include weekly retail short sell share of volumes [t-5, t-1] *rtlshort_w*, interactions between *rtlshort_w* and *news*, dummy variable *news*(t=0). I also include previous week return [t-5, t-1] *ret(w-1)*, previous month returns [t-25, t-6] *ret (m-1)*, previous year returns [t-251, t-26] *ret(y-1)*. I also include log market cap (*size*), log book to market ratio (*btm*), monthly turnover (*turn*), and monthly volatility of daily returns (*vol*), all measured at the end of previous month. To account for serial correlation in the coefficients, the standard deviations of time-series are adjusted using Newey-west with 5 lags and 20 lags. Dependent variables are in percent, and independent variables except dummy variable *news* are standardized each day t.

Panel A: All the news

	Y=Cumulative returns [t+1, t+5]			Y=Cumulative returns [t+1, t+20]		
	1	2	3	4	5	6
Rtlshort_w	-0.018***	-0.016***	-0.016***	-0.063***	-0.062***	-0.061***
	(-3.01)	(-2.75)	(-2.71)	(-3.02)	(-2.99)	(-2.94)
Rtlshort_w*News		-0.018***	-0.015***		-0.015	-0.018*
		(-3.71)	(-2.88)		(-1.21)	(-1.65)
News			-0.031			-0.021
			(-1.22)			(-0.28)
Controls	yes	yes	yes	yes	yes	yes
R square	0.044	0.044	0.045	0.041	0.042	0.042
# of days	1736	1736	1736	1736	1736	1736
Obs	5540550	5540550	5540550	5540550	5540550	5540550

Panel B: Negative news

	Y=Cumulative returns [t+1, t+5]			Y=Cumulative returns [t+1, t+20]		
	1	2	3	4	5	6
Rtlshort_w	-0.018***	-0.014**	-0.016***	-0.060**	-0.054**	-0.057**
	(-2.89)	(-2.15)	(-2.61)	(-2.57)	(-2.35)	(-2.46)
Rtlshort_w*Negative news		-0.060***	-0.017***		-0.065***	-0.027**
		(-11.89)	(-3.18)		(-4.81)	(-2.22)
Negative news			-0.308***			-0.311***
			(-8.78)			(-3.17)
Controls	yes	yes	yes	yes	yes	yes
R square	0.046	0.047	0.047	0.043	0.044	0.044
# of days	1736	1736	1736	1736	1736	1736
Obs	5135789	5135789	5135789	5135789	5135789	5135789

Table 7: Cross-sectional Regressions of Returns on Retail Trading and Retail Short Selling: News vs. Non-News

This table presents daily Fama Macbeth (1973) regressions of returns on retail short selling and retail order imbalances, by dividing days into news days and non-news days. Sample period is from January 2010 to December 2016. The news dummy equals to 1 if there is earning announcement, or analyst recommendation change, or analyst forecast revision on that day t=0, else equals to zero. The dependent variables are cumulative raw returns over [t+1, t+5]. Independent variables include total weekly retail order imbalance [t-5, t-1] (*rtloib_total_w*), weekly retail short sells [t-5, t-1] (*rtlshort_w*), weekly retail order imbalance [t-5, t-1] (*rtloib_w)*, and their interactions with dummy variable *news*, as well as *news*(t=0). Independent variables also include previous week return [t-5, t-1] *ret(w-1)*, previous month returns [t-25, t-6] *ret (m-1)*, previous year returns [t-251, t-26] *ret(y-1).* I also control log market cap (*size*), log book to market ratio (*btm*), monthly turnover (*turn*), and monthly volatility of daily returns (*vol*), all measured at the end of previous month. To account for serial correlation in the coefficients, the standard deviations of time-series are adjusted using Newey-west with 5 lags. Dependent variables are in percent, and independent variables except dummy variable *news* are standardized each day t.

	Y=Cumulative returns [t+1, t+5]	
	1	2
Rtloib_total_w	0.053***	
	(12.67)	
Rtloib_total_w*News	-0.001	
	(-0.26)	
Rtlshort_w		-0.015***
		(-2.59)
Rtlshort_w*News		-0.018***
		(-2.58)
Rtloib_w		0.052***
		(12.07)
Rtloib_w*News		-0.004
		(-0.74)
News	0.021	0.048*
	(0.89)	(1.92)
Controls	yes	yes
R square	0.045	0.047
# of days	1736	1736
Obs	5540550	5540550

Table 8: Cross-sectional Regressions of Determinants of Retail Short Selling

This table presents daily Fama Macbeth (1973) regressions of determinants for retail short sells. Sample period is from January 2010 to December 2016. Dependent variable is retail short sell percentage of volumes on day t=0 *rtlshort*, and independent variables include contemporaneous day return t=0 *ret(0)*, previous week return [t-5, t-1] *ret(w-1)*, previous month returns [t-25, t-6] *ret (m-1)*. I also include log market cap (*size*), log book to market ratio (*btm*), monthly turnover (*turn*), and monthly volatility of daily returns (*vol*), all measured at the end of previous month. In addition, I also include weekly retail short sell percentage of volumes [t-5, t-1] *rtlshort_w*, and the most recent short interest *shortint*, which can be observed on day t=0. To account for serial correlation in the coefficients, the standard deviations of time-series are adjusted using Newey-west with 5 lags. All the dependent variable and independent variables are standardized each day t.

	Y=Rtlshort [t=0]		
	1	2	3
Rtlshort_w			0.106***
			(47.27)
Shortint			-0.002
			(-1.43)
Log(Size)		-0.074***	-0.059***
		(-12.79)	(-12.98)
Log(Btm)		-0.018***	-0.015***
		(-19.08)	(-21.50)
Turn		0.015***	0.012***
		(9.09)	(10.22)
Vol		0.013***	0.011***
		(6.58)	(6.86)
Ret (0)	0.014***	0.014***	0.014***
	(13.61)	(15.28)	(15.39)
Ret (w-1)	0.019***	0.019***	0.017***
	(14.21)	(17.98)	(18.45)
Ret (m-1)	0.011***	0.010***	0.007***
	(6.50)	(8.83)	(8.21)
Intercept	-0.000	-0.003***	0.000***
	(-0.43)	(-8.57)	(10.08)
R square	0.003	0.014	0.027
# of days	1736	1736	1736
Obs	5542587	5542587	5542587

Table 9: Cross-sectional Regressions of Retail Short Selling with Buying Pressure

This table presents daily Fama Macbeth (1973) regressions of retail short sells with buying pressures on individual stocks. Sample period is from January 2010 to December 2016. Dependent variable is retail short sell share of volumes on day t=0 *rtlshort*. Independent variables include positive contemporaneous Lee and Ready order imbalance *LRoib+(0)* and weekly Lee and Ready order imbalance [t-5, t-1] *LRoib+_w*. LRoib is calculated as (buy volumes-sell volumes)/total volumes, and LRoib+ equals LRoib if LRoib>0 and zero otherwise. Independent variables also include positive contemporaneous retail order imbalance *rtloib+(0)*, weekly retail order imbalance [t-5, t-1] *rtloib+_w*, positive contemporaneous other order imbalance *otheroib+* and weekly other order imbalance [t-5, t-1] *otheroib+_w* . Rtloib is calculated as (retail buy volumes – retail long sell volumes)/total volumes, and Rtloib+ equals rtloib if rtloib>0 and zero otherwise. Otheroib is calculated as ((buy volumes-retail buy volumes)-(sell volumes-retail sell volumes))/total volumes, and otheroib+ equals otheroib if otheroib>0 and zero otherwise. Other control variables include contemporaneous day return t=0 *ret(0)*, previous week return [t-5, t-1] *ret(w-1)*, previous month returns [t-25, t-6] *ret (m-1)*, weekly retail short sell share of volumes [t-5, t-1] *rtlshort_w*, and the most recent short interest *shortint*, which can be observed on day t=0. I also include log market cap (*size*), log book to market ratio (*btm*), monthly turnover (*turn*), and monthly volatility of daily returns (*vol*), all measured at the end of previous month. To account for serial correlation in the coefficients, the standard deviations of time-series are adjusted using Newey-west with 5 lags. All the dependent variable and independent variables, except order imbalance, are standardized each day t. Order imbalance is not demeaned but standardized to have unit standard deviation before partitioning into positive and negative values.

	Y=Rtlshort [t=0]			
	1	2	3	4
LRoib+	0.014***			
	(4.21)			
LRoib+_w	0.010***			
	(10.29)			
Rtloib+		0.021***		0.025***
		(7.31)		(9.16)
Rtloib+_w		0.032***		0.029***
		(34.90)		(35.09)
Otheroib+			0.153***	0.149***
			(27.84)	(27.57)
Otheroib+_w			-0.008***	-0.010***
			(-2.00)	(-11.34)
Controls	yes	yes	yes	yes
R square	0.030	0.033	0.040	0.045
# of days	1736	1736	1736	1736
Obs	5542587	5542587	5542587	5542587

Table 10: Cross-sectional Regressions of Retail Short Selling on Returns under Different Conditions

This table presents daily Fama Macbeth (1973) regressions of retail short sells on returns and other control variables on the stocks with or without news in Panel A and with or without buying pressure in Panel B. Sample period is from January 2010 to December 2016. Dependent variable is retail short sell share of volumes on day t=0 *rtlshort*. In Panel A, news dummy equals to one if there is earning announcement, or analyst recommendation change, or analyst forecast revision on that day t=0, else equals to zero. I include interaction terms of contemporaneous and past returns with news dummy, returns, and news dummy. In Panel B, order imbalance LRoib+ /rtloib+/otheroib+ equals LRoib /rtloib /otheroib if greater than zero, and otherwise zero. I include interaction terms of contemporaneous and past returns with non-negative order imbalance, returns and non-negative order imbalance. To account for serial correlation in the coefficients, the standard deviations of time-series are adjusted using Newey-west with 5 lags. All the dependent variables and independent variables except order imbalance and news, are standardized each day t. Order imbalance is not demeaned but standardized to have unit standard deviation before dividing into positive and negative values.

Panel A: Interacted with news days

	Y=Rtlshort [t=0]		
	1	2	3
Ret (0)*News	-0.006***		
	(-8.82)		
Ret (w-1)*News	-0.002***		
	(-5.78)		
Ret (m-1)*News	-0.001		
	(-1.55)		
Ret (0)*Negative news		-0.009***	
		(-10.78)	
Ret (w-1)*Negative news		-0.004***	
		(-8.35)	
Ret (m-1)*Negative news		-0.001**	
		(-2.42)	
Ret (0)*Positive news			-0.000
			(-0.67)
Ret (w-1)*Positive news			0.000
			(0.30)
Ret (m-1)*Positive news			-0.000
			(-0.06)
Ret (0)	0.016***	0.017***	0.014***
	(16.25)	(15.66)	(12.23)
Ret (w-1)	0.019***	0.021***	0.020***
	(14.76)	(13.70)	(13.16)
Ret (m-1)	0.011***	0.010***	0.011***
	(6.69)	(5.57)	(5.69)
News (0)	-0.005	-0.008	-0.010*
	(-0.98)	(-1.42)	(-1.83)
Intercept	0.001*	0.000*	0.001***
	(1.67)	(1.80)	(2.65)

R square	0.003	0.004	0.004
# of days	1736	1736	1736
Obs	5540542	5136506	5120716

Panel B: Interacted with buying pressure

	Y=Rtlshort [t=0]		
	1	2	3
Ret (0)*LRoib+	0.014***		
	(10.42)		
Ret (w-1)*LRoib+	0.004***		
	(3.20)		
Ret (m-1)*LRoib+	0.003**		
	(2.33)		
Ret (0)*Rtloib+		0.003**	
		(2.49)	
Ret (w-1)*Rtloib+		0.010***	
		(8.98)	
Ret (m-1)*Rtloib+		0.009***	
		(8.46)	
Ret (0)*Otheroib+			0.010***
			(5.23)
Ret (w-1)*Otheroib+			-0.000
			(-0.03)
Ret (m-1)*Otheroib+			0.001
			(0.44)
Ret (0)	0.004***	0.011***	-0.003**
	(3.01)	(9.14)	(-2.21)
Ret (w-1)	0.017***	0.016***	0.016***
	(12.76)	(12.21)	(12.50)
Ret (m-1)	0.010***	0.008***	0.010***
	(5.53)	(5.20)	(5.96)
Oib+(0)	0.041***	0.062***	0.159***
	(9.16)	(17.85)	(22.55)
Intercept	-0.013***	-0.013***	-0.049***
	(-8.90)	(-17.00)	(-22.50)
R square	0.010	0.009	0.027
# of days	1736	1736	1736
Obs	5542587	5542587	5542587

Appendix A1: Portfolio Returns on days [x,y] based on Retail Short Selling

This table presents daily value-weighted returns for portfolios based on weekly retail short sells from January 2010 to December 2016. Each day, stocks are sorted into quintiles based on weekly retail short selling's percentage of volumes (*rtlshort_w*). After skipping 1 day, value-weighted portfolios are held for 5 trading days [1,5], 20 trading days [1,20], 40 trading days, and 60 trading days, respectively. This process is repeated each trading day, so each trading day's portfolio return with horizon [x,y] is an average of (y-x+1) different portfolios, with 1/(y-x+1) of the portfolio rebalanced each day, following Jegadeesh and Titman (1993). Daily calendar time returns and Fama and French (1993) three factor alphas are reported in percent. Newey-west (1987) t-statistics based on five lags appear in parentheses.

	Retail shorting's percentage of volume (*rtlshort_ w*)							
	[1,5]		[1,20]		[21,40]		[41,60]	
	Return	Alpha	Return	Alpha	Return	Alpha	Return	Alpha
1 (Least)	0.064	0.019	0.061	0.017	0.065	0.014	0.062	0.015
2	0.062	0.012	0.059	0.010	0.065	0.009	0.059	0.006
3	0.054	0.004	0.050	0.002	0.056	0.001	0.056	0.003
4	0.044	-0.008	0.045	-0.006	0.051	-0.007	0.047	-0.008
5(Most)	0.048	-0.009	0.049	-0.006	0.059	-0.003	0.058	0.001
5-1	-0.016	-0.028***	-0.012	-0.024***	-0.006	-0.017*	-0.004	-0.015
	(-1.362)	(-2.695)	(-1.185)	(-2.606)	(-0.603)	(-1.783)	(-0.398)	(-1.540)

Appendix A2: Cross-sectional Regressions of Returns on Retail Short Selling

This table presents daily Fama Macbeth (1973) regressions of next week's return [t+1, t+5] on weekly retail short selling [t-5, t-1] (*rtlshort_w)* and control variables measured as of day t. Sample period is from January 2010 to December 2016. Dependent variable in Panel A is cumulative raw returns [t+1, t+5], and dependent variable in Panel B is cumulative risk-adjusted returns [t+1, t+5] using Fama-French three factor model, with betas estimated from previous year. I also include weekly dark pool non-retail short sells [t-5, t-1] (*darkshort_w)*, calculated as (OTC short sell volumes-retail short sell volumes)/total volumes, and weekly retail order imbalance [t-5, t-1] (*rtloib_w)*, calculated as (retail buy volumes-retail *long sell* volumes)/total volumes. Other control variables include previous week return [t-5, t-1] *ret(w-1)*, previous month returns [t-25, t-6] *ret (m-1)*, and previous year returns [t-251, t-26] *ret(y-1)*. I also include log market cap (*size*), log book to market ratio (*btm*), monthly turnover (*turn*), and monthly volatility of daily returns (*vol*), all measured at the end of previous month. To account for serial correlation in the coefficients, the standard deviation of time-series is adjusted using Newey-west with 5 lags. Dependent variables are in percent, and all the independent variables are standardized each day t.

Panel A:

	Y= Cumulative raw returns [t+1, t+5]			
	1	2	3	4
Rtlshort_w	-0.031***	-0.018***	-0.017***	-0.022***
	(-4.13)	(-3.10)	(-2.99)	(-3.78)
Darkshort_w			-0.009	
			(-1.01)	
Rtloib_w				0.050***
				(11.53)
Log(Size)		0.038*	0.037*	0.038*
		(1.79)	(1.74)	(1.76)
Log(Btm)		0.026	0.025	0.026
		(1.50)	(1.46)	(1.54)
Turn		-0.049***	-0.047***	-0.050***
		(-3.48)	(-3.50)	(-3.53)
Vol		-0.090***	-0.089***	-0.089***
		(-4.48)	(-4.50)	(-4.47)
Ret(w-1)		-0.091***	-0.090***	-0.093***
		(-6.16)	(-6.07)	(-6.31)
Ret (m-1)		-0.011	-0.011	-0.010
		(-0.66)	(-0.67)	(-0.58)
Ret (y-1)		0.053**	0.053**	0.053**
		(2.54)	(2.54)	(2.52)
Intercept	0.241**	0.241**	0.241**	0.241**
	(2.25)	(2.25)	(2.25)	(2.25)
R square	0.001	0.043	0.045	0.044
# of days	1736	1736	1736	1736
obs	5542587	5542587	5542587	5542587

Panel B:

	Y= Cumulative risk-adjusted returns [t+1, t+5]			
	1	2	3	4
Rtlshort_w	-0.024***	-0.015***	-0.013***	-0.019***
	(-3.35)	(-2.85)	(-2.67)	(-3.60)
Darkshort_w			-0.015*	
			(-1.94)	
Rtloib_w				0.048***
				(11.47)
Log(Size)		-0.037**	-0.038**	-0.037**
		(-2.50)	(-2.53)	(-2.52)
Log(Btm)		0.034***	0.033***	0.034***
		(3.57)	(3.42)	(3.62)
Turn		-0.053***	-0.052***	-0.054***
		(-3.94)	(-3.91)	(-3.99)
Vol		-0.091***	-0.090***	-0.091***
		(-6.11)	(-6.09)	(-6.10)
Ret(w-1)		-0.088***	-0.087***	-0.090***
		(-7.14)	(-6.99)	(-7.31)
Ret (m-1)		-0.004	-0.004	-0.002
		(-0.26)	(-0.27)	(-0.17)
Ret (y-1)		0.051***	0.050***	0.050***
		(2.99)	(2.96)	(2.96)
Intercept	0.017	0.016	0.016	0.016
	(1.29)	(1.28)	(1.27)	(1.28)
R square	0.001	0.027	0.028	0.028
# of days	1736	1736	1736	1736
obs	5364476	5364476	5364476	5364476

Chapter 3

Passive Investing, Stock Price Efficiency and Liquidity

Abstract

This paper studies the market quality of stocks that are owned by funds using passive investment strategies (indexers). We find that indexing significantly improves the underlying stocks' market quality, as reflected in increased liquidity and price efficiency. First, consistent with an arbitrage hypothesis, more trading in the index security itself allows arbitrageurs to help spread systematic information from indexes to underlying securities. When we compare ETFs to index funds, we find that greater ETF holdings, but not greater index fund holdings, increase the volatility of underlying stocks. Second, consistent with a short selling hypothesis, stock addition into indexes increases the available lendable shares and allows short sellers to express negative news, which correct overpriced hard-to-borrow stocks. Thus, indexing improves market quality in the underlying stock, but ETF ownership partly reverses this effect.

3.1. Introduction

Indexing allows investors to diversify at low costs. The sector as a whole has expanded tremendously and at increasing rates over the past 25 years. From 2007 to 2016 alone, indexing by domestic equity mutual funds and by exchange-traded index funds (ETFs) has generated $1.4 trillion in net new cash and reinvested dividends[9]. Given the large size of indexing industry, their impact on equity markets becomes an important concern on investors, researchers, and regulators.

The growth of indexing is consistent with French (2008), who argues that passive investing is beneficial to investors and states that "the typical investor would increase his average annual returns by 67 basis points over the 1980-2006 period if he switched to a passive market portfolio." Stambaugh (2014) relates the increase in passive investing to a decline in noise trading, which benefits the market as a whole. Many researchers, however, do not share a positive view of indexing. Wurgler (2011) warns about adverse effects of increasing indexing. In particular, he argues that more indexing can generate excess co-movement (Da and Shive, 2016) and higher volatility (Ben-David, et al. 2018). We contribute to this discussion and examine the role of indexing on underlying market quality and the efficiency of markets that trade underlying stocks.

Using cross-sectional index holding data from 2002 to 2016, we measure index ownership by the percentage of shares held by passive index funds at the end of each quarter. There are 698 passive indexers in our sample, including 355 ETFs and 343 index mutual funds. Index ownership as whole rises from less than 2% in 2002 to over 8% in 2016, with a mean of 4.9%.

Variation in firm-level passive holdings can come from two reasons: from purchase and redemption activities at the end of the day, perhaps in response to excess demand or supply from investors; or from regular index rebalancing, which forces index funds to update their portfolio.

[1] According to Morningstar Inc, in 2016, passive funds in the United States attracted $506 billion, and actively managed funds posted $341 billion in withdrawals.

Therefore, we use changes in index ownership as total passive investing,[10] and construct stock-level proxies for both sources of changes in passive portfolios.

In our first set of tests, we highlight the effects of passive investing, including index flows and index rebalancing, on liquidity and price efficiency in the market for the underlying stocks. We find that more indexing is associated with greater liquidity and more efficient prices in the underlying stocks. These results control for institutional trading, analyst following, and stock level characteristics such as market cap, book to market ratio, stock price, turnover and volatility, and are robust to several alternative model specifications. Moreover, increases in the index and inflows both improve liquidity, but inflows, rather than changes in the index, are the main reasons for the positive effects that are associated with increases in ownership by passive investors.

We measure price efficiency in three different ways. First, we use annual and quarterly price delay based on Hou and Moskowitz (2005). The delay captures the portion of individual stock return variation that can be explained by (observable) lagged market returns. Our results show that price delay decreases with increasing passive investing, indicating that indexing makes the incorporation of market-wide information faster.

Second, under the assumption that efficient prices follow a random walk, the deviation of stock prices from a random walk is a measure of relative inefficiency. To measure the deviation from a random walk we use variance ratios. Specifically, we compute the deviation of the weekly return variance relative to five times the daily return variance. If prices follow a random walk, then this ratio should equal one. (Lo and MacKinlay, 1988). We find that only stock-level fund flows decrease the variance ratios and thus increases underlying price efficiency. In contrast, index trading caused by index rebalancing instead decreases the price efficiency of the component stocks.

[2] To improve our ability to identify the consequences of passive investing in a clean setting, we focus on the effects of lagged changes of passive ownership.

Third, we differentiate changes in price efficiency that are due to systematic information from those changes that are due to firm-specific information. We use market returns and industry returns to construct a return synchronicity measure (Crawford, Roulstone, and So, 2012), and find that index trading significantly increases the proportion of systematic information that is integrated into stock prices.

Our main conclusion is that indexing improves the market quality of the index's underlying stocks. Market quality improves in terms of greater liquidity, more return co-movement, and better price efficiency. By dividing passive investing into index fund flows and index rebalancing, we document their relative impact on underlying stocks. On one hand, Cong and Xu (2016) model passive indexes as common factor investing. Consistent with the literature, our result shows that passive fund flows significantly improve liquidity and price efficiency of the component stocks, which attract both liquidity traders who attempt to achieve diversification with low adverse selection and informed investors who prefer to trade on their macro-based information through low-cost passive funds. On the other hand, Bessenbinder (2015) attributes stock additions or deletions into index funds and ETFs to predictable institutional order flows. These predictable orders would have minimal effects on prices but could benefit in liquidity supply. Accordingly, we provide evidence that index rebalancing indeed improves underlying stocks' liquidity. Index rebalancing cannot be attributed to firm fundamentals, but it offers greater opportunities for liquidity providers.

In the second part of our paper we explain the channels through which passive investing impacts underlying market quality. The first is an arbitrage argument. The idea is that indexers allow arbitrageurs to establish positions that benefit from price discrepancies between indexes and the constituent stocks (Fremault, 1991; Kumar and Seppi, 1994). We conjecture that greater passive investing reduces information asymmetry and order imbalances, and thus improves liquidity and lowers arbitrage costs and arbitrage risk. As such, it is conceivable that index prices reflect an aggregate version of news first, before it is incorporated into underlying prices. Arbitrageurs who are active in index products help both prices and NAVs to adjust and allow systematic information flow from passive funds to underlying securities, causing a closer link between fundamentals and stock prices.

To test the arbitrage channel, we first construct aggregate passive fund flows each period, motivated by Akbas, et al. (2015). If index trading contains information, greater aggregate fund flows would contribute to more efficient prices by facilitating arbitrage activities and correcting mispricing. We thus interact period dummy variables with passive investing, and find that greater passive investing significantly decreases arbitrage risk and increases stock price efficiency during high aggregate flow periods. Especially, index trading represented by stock-level index fund flows only improves underlying price efficiency during the high aggregate flow periods. Second, compared with traditional open-end index mutual funds, which are traded at net asset values at the end of day, ETFs can be traded throughout the day, which causes ETF prices in the secondary market to deviate from net asset values at any time and allows market makers to simultaneously take opposite positions in ETFs and underlying shares. Therefore, we separately consider the effects of ETFs and effects of index mutual funds on underlying market quality. We find that, it is the ETFs that increase underlying volatility and variance ratio, which may partially reverse the effects.

Secondly, we posit a short-selling story to explain the channel through which passive investing impacts the component stocks' market quality. Nagel (2005) finds that stocks with greater passive holdings have larger stock lending supplies, which can partially mitigate the under-performance caused by short sell constraints. Prado, Saffi, and Sturgess (2016) further indicate that stocks with high index ownership are associated with higher lending supply, lower borrowing costs and lower arbitrage risk. Thus, we assume that if a stock is added to an index, while not being dropped from another, the number of shares available for shorting goes up, possibly reducing the shorting costs and reducing the cost of a negative bet. Therefore, greater passive investing, especially the trading related with index rebalancing, would relax short selling constraints and thus improve liquidity and price efficiency by reducing the costs of expressing negative views through short selling (Boehmer and Wu, 2013; Saffi and Sigurdsson, 2011).

To test the short selling channel, we first find that greater passive holding is related with more short sales. In passive trading, the main increase in short selling is from index rebalancing rather than from index flows. Besides, we isolate positive market returns and negative market returns to reconstruct price delay. Result shows that greater indexing would increase the speed of negative information incorporated into stock prices by facilitating greater short selling activities. This effect especially focuses on the stocks which were hard to borrow [11]under market. Our result is consistent with the findings in Blocher and Whaley (2016) that ETF managers respond to the lending incentives by slanting their holdings to the hard to borrow stocks.

Finally, in this paper, we also consider the difference in underlying stocks. Glosten, Nallareddy and Zou (2017) document that ETF activity increases information efficiency for stocks with weak information environments. These stocks have less public available information and greater informational asymmetry, and thus arbitrage is more constrained and short selling is relatively limited. But passive indexes provide a low-cost trading avenue for both arbitrage activities and loan supplies, which will increase liquidity and price efficiency more for these stocks. We sort our sample of stocks by market cap and liquidity, and find that the increase in market quality brought by passive investing is mainly concentrated in stocks of smaller size and lower

[11] Vanguard has designed its securities' lending program to capture the scarcity premium found in many hard to borrow stocks.

turnover. Especially, when we use stock-level fund flows to measure index trading, the increased liquidity and price efficiency are only concentrated on the small and illiquid stocks. In contrast, large-cap stocks have even less efficient prices with greater index fund flows.

Our study first contributes to a growing literature on the economic consequences of basket or index products. Adopting key insights from information economics, we present empirical evidence on how incentives in the market for information can affect pricing for the underlying securities. In the literature, Subramanyam (1991) and Gorton and Pennacchi (1993) propose a ‘Substitution Hypothesis’ that uninformed investors will migrate to indexes because their losses to informed traders are lower in this market, which generates a disincentive for informed traders to expend resources for firm-specific information. Later studies (Hamm 2014; Qin and Singal, 2015; Israeli, Lee, and Sridharan, 2017) are built on this story and find that passive indexes including ETFs decrease underlying liquidity and price efficiency. But in their model, only liquidity noise traders are attractive to indexes. We argue that the liquidity trading in passive indexes can also attract other investors to trade on information. Our paper is therefore from an ‘Arbitrage Hypothesis’ (Fremault, 1991; Kumar and Seppi, 1994) and contributes to the debate about whether arbitrage activities transmit only liquidity shocks (Da and Shive, 2016; Ben-David, Franzoni, Moussawi, 2018; B’Hattacharaya and O’Hara, 2018) or fundamental information (Hasbrouck, 2003; Boehmer and Boehmer, 2003; Cong and Xu, 2016; Glosten, Nallareddy, and Zou, 2017). We provide evidence about positive impacts of passive investing and indicate that arbitrage activities are effective, which indeed transmit systematic information to the underlying stocks, and benefit in greater liquidity and better price efficiency.

In addition, our paper complements the study about loan activities brought by passive indexes. Active shareholders are less likely to lend shares on a large scale since the ownership and voting rights will be transferred and the lack of voting rights is known to discourage the participation of active institutional investors (Nagel, 2005; Prado, et al.2016; Evans, et al. 2017). In comparison, passive funds are more likely to provide lendable shares and earn lending fees (Blocher and Whaley, 2016). We accordingly provide evidence that passive investing indeed promotes short selling activities. Especially, index addition increases the proportion of available lendable shares. By relaxing short sale constraints and reducing borrowing costs, passive investing increases the speed of negative news incorporated into stock prices.

Finally, this paper compares the differences between index mutual funds and ETFs. Compared with index mutual funds, ETFs allow intraday trading on stock exchanges. Ben-David, et al (2018) find that ETF holding increases volatility and transmits liquidity shocks to underlying stocks, which will decrease underlying market quality. We find that, it is the ETFs, that increase underlying volatility, which may partially reverse the positive effect from index mutual funds. But overall, index mutual funds will increase underlying price efficiency, by helping transmit systematic information through greater arbitrage activities and by relaxing short selling constraints through more loan supplies.

The paper is organized as follows. Section 2 describes data sources, sample construction, and variable measures. Section 3 presents main empirical results of this paper, including stock-level regression results of passive investing on stock liquidity and price efficiency. Section 4 tests three related channel attempting to explain the relations. Section 5 provides robust tests, including propensity score matching to resolve causality issues and Section 6 gives conclusions.

3.2.Data and Sample construction

3.2.1. Sample Construction

Index mutual funds and ETFs are the major passive investing vehicles in the market. We obtain passive index data on domestic equity mutual funds from the CRSP mutual fund database, selecting all funds that are classified as either index funds or ETFs by CRSP index fund/ETF indicators. To identify and recover passive funds that are not marked by indicators, we screen the remaining sample on keywords in their names. A fund is classified as passive if it calls itself "index" or "ETF" [12](Qin and Signal, 2015). We exclude bond funds and international funds. Additionally, we require that equity funds in our sample hold between 80% to 105% of their portfolio in common stocks. Moreover, funds must hold at least 10 stocks and manage assets of more than 5 million USD. Since fund characteristics provided by CRSP are at the share class level, we calculate value-weighted fund characteristics across multiple share classes within an index using TNA as weights, except that TNA is the sum of net assets across all share classes belonging to a given fund. The above procedure yields a final sample of 698 passive funds from 2002.q1 to 2016.q4.

Next, quarterly index holdings data are from Thomson-Reuters' Mutual Fund holding database (S12), which we merge with total indexes sample using MFLINKS. Index ownership for each stock $i$at the end of quarter t is calculated as:

$$Index\%_{i,t} = \frac{\sum_{j=1}^{J} shares_{i,j,t}}{shares\ outstanding_{i,t}} \tag{1}$$

where J is the set of index funds holding stock i; $shares_{i,j,t}$ is the number of stock i shares held by index j at the end of quarter t; $shares\ outstanding_{i,t}$ is the total shares outstanding for stock i at the end of quarter t. We only consider common stocks that have share codes 10 or 11 traded on three main exchanges. Stocks in the sample have prices between \$1 and \$1000, and have available data for share price, shares outstanding, and book value of equity in each quarter. We further require that stocks in the sample have at least positive index holding and institutional

[12] For index funds, we search the following keywords: 'INDEX', 'IND', 'IDX', INDX', 'S&P 500 I', 'S&P 500I', 'S&P 400 I', 'S&P 400I', 'S&P 600I', 'S&P500IND', 'S&P400IND', 'S&P600IND', 'RUSSELL 1000', 'RUSSELL2000', 'RUSSELL 3000', "NASDAQ','NYSE' and 'VANGUARD'. For ETFs, we search the following keywords: 'EXCHANGE TRADED', 'EXCHANGE-TRADED', 'ETF', 'ISHARES', 'POWERSHARES', 'PROFUNDS', 'SPDR S&P', 'SPDR DOW', 'SPDR DJ, 'RYDEX', 'SPA MG', 'MARKET GRADER', and 'QQQ'.

holding. The final sample on average covers 3,453 firms each quarter and 183,570 firm-quarters. Figure 1 presents the average index ownership across firms each year. There is a significant increase in average index ownership from less than 2% in 2002 to more than 8% in 2016, with a mean level of around 4.88%

3.2.2. Passive investing proxies

Along with increasing passive ownership, there is greater passive investing. We take change of passive ownership as total passive trading on the individual stock, which can be further divided into two parts.

One is caused by fund flows from investors, represented by nightly subscription and redemption process at the net asset values. As more investors subscribe to the fund, its assets would increase in response to fund flows. Therefore, funds can grow or shrink based on net investor demands at end of day. The investment flow to fund j in quarter t is calculated as (Sirri and Tuffano, 1998):

$$Flow_{j,t} = \frac{TNA_{j,t} - TNA_{j,t-1} * (1 + qret)}{TNA_{j,t-1}} \tag{2}$$

We assume that inflows and outflows occur at the end of each quarter. Mutual funds that are initiated have inflows equal to their initial TNA, while funds that are liquidated have outflows equal to their terminal TNA. In response to inflows or outflows, index funds expand or liquidate exiting holdings proportionally, rather than disproportionally on some securities as active mutual funds do. So, by using lagged index ownership as a weight to average index fund flows across all passive funds holding the stock, flow-induced trading for underlying stock i in quarter t is as:

$$Stock\_flow_{i,t} = \frac{\sum_{j=1}^{J} Index\%_{i,j,t-1} * Flow_{j,t}}{\sum_{j=1}^{J} Index\%_{i,j,t-1}} \tag{3}$$

The other part causing the change of passive ownership is from deviation investing such as index rebalancing. We extract this part from total passive trading using the following regression:

$$\Delta Index\%_{i,t} = \beta_{0,i} + \beta_{i,i} Stock\_flow_{i,t} + \eta_t + \varepsilon_{i,t} \quad (4)$$

In the regression, we control quarter fixed effects and cluster standard errors on firm level. The residual part in the regression, $flow\_residual$, is taken to proxy for the impacts of index rebalancing on underlying stocks.

3.2.3 Liquidity measures and Price efficiency measures

We use two measures to calculate stocks liquidity. We compute Amihud's (2002) ratio as absolute daily returns over daily dollar volume in millions and bid-ask spread as difference between daily bid and ask price divided by daily closed price to proxy for stock illiquidity. Data are from the CRSP/daily stock file. Since our regression is at a quarterly frequency, we average the daily variable over each quarter. On average, stock has mean level of spread at 0.7% and Amihud ratio of 1.0554.

We employ three different approaches to measure price efficiency. First, Hou and Moskowitz (2005) introduce a "price delay" that relies on the speed of adjustment to market-wide information. We replicate their annual delay measure and, additionally, create an analogous quarterly measure (Boehmer and Wu, 2013). We require a minimum of 50 observations per firm-quarter, using daily rather than weekly observations. We regress stock returns on contemporaneous and five days of lagged market returns over each quarter and calculate R^2 of the regression:

$$r_{i,t} = \alpha_i + \beta_i^0 R_{m,t} + \sum_{n=1}^{5} \beta_i^n R_{m,t-n} + \varepsilon_{i,t} \quad (5)$$

Next, we estimate a restricted model that limits the coefficients of lagged market returns to zero and also acquire R^2. Then the delay measure by combining the restricted model and the unrestricted model R^2 is as:

$$Delay\_r = 1 - \frac{R^2_{restricted}}{R^2_{unrestricted}} \quad (6)$$

The larger this measure, the greater the variation in stock returns explained by lagged market returns, which implies a longer price delay in response to market information. However, this measure doesn't take the magnitude of lagged market returns' coefficients into account. Therefore, we similarly construct a delay measure based on the regression coefficients (Saffi and Sigurdsson, 2011) as:

$$Delay\_coef = \frac{\sum_{n=1}^{5} n * \frac{abs(\beta_i^n)}{se(\beta_i^n)}}{\frac{abs(\beta_i^0)}{se(\beta_i^0)} + \sum_{n=1}^{5} \frac{abs(\beta_i^n)}{se(\beta_i^n)}} \tag{7}$$

This measure captures the magnitude of lagged coefficients relative to magnitude of all market return coefficients. Greater delay indicate that price efficiency is smaller. On average, in our sample, $Delay\_r$ averages at 0.3819 and $delay\_coef$ averages at 1.6679.

Second, under the assumption that the efficient price follows a random walk process, we can measure price efficiency as the deviations of stock prices from a random. We use the deviation of variance ratios of weekly to daily returns from one (Lo and MacKinlay, 1988). For each stock, we calculate the absolute deviation of the ratio of the weekly return variance to five times the daily return variance from one during a quarter, $|1 - VR\ (1,5)|$, where the weekly returns are estimated from Wednesday to Wednesday to eliminate the weekend effect. A large variance ratio means that stock prices have greater deviation from a random walk and thus are less efficient. For quarterly regressions, we average this measure within quarters. A stock has averaged variance ratio of 0.3376.

Finally, to measure the systematic information incorporated into stock prices, we determine the degree of co-movement of individual stock returns with market returns and industry returns (Crawford, Roulstone, and So, 2012). We require that there are at least 50 observations per firm-quarter to run the regression. For each firm-quarter, we regress daily stock excess return on market excess return to obtain $beta$ and R^2. Thus, return synchronicity using market returns is as:

$$Sync\_mkt_{i,t} = \log\left(\frac{R_{i,t}^2}{1-R_{i,t}^2}\right) \tag{8}$$

$Sync\_ind$ is similarly defined, but also includes industry returns in the regression. Piotroski and Roulstone (2004) explain synchronicity changes as whenever new information alters investors' understanding of how the fundamentals of the firm align with the fundamentals of the industry or overall markets. Therefore, high value of synchronicity indicates that a greater fraction of firm-level return variation is explained by general market and related-industry return variation.

3.2.4 Other Control Variables

We choose control variables related to firm's informational environment, beginning with firm size, book-to-market, stock price, share turnover and return volatility. Size is price*shares outstanding from CRSP, calculated daily and averaged over each quarter. Btm is calculated as book value for the fiscal year ended before the most recent June 30, divided by market capitalization of December 31 during that fiscal year and the data are from Compustat. Turnover is the ratio of daily trading volume to total shares outstanding. Stock prices are closing share prices at the end of the quarter. We limit the sample to stocks with prices between $1 and $1000. Volatility is proxied by standard deviation of daily stock returns during the quarter.

We include institutional ownership since stocks with greater institutional ownership are more liquid and priced more efficiently (Boehmer and Kelley, 2009). We calculate the ratio as shares outstanding held by institutional investors in Thomson Reuters/13f fillings. We exclude sample stocks with institutional ownership over 100% or stocks without institutional holdings. Stocks in our sample have institutional ownership of 53.46% and are held by 130 large institutions. We use the number of analysts following a stock as a proxy for information production. Fewer analysts following means that firms are active in relatively poor information environments. Analyst is measured as the total number of analysts that report earning forecasts for a stock from

I/B/E/S unadjusted detail data. If data is missing, we assign number of analysts to zero. On average, a stock is followed by #7.1 analysts at the same time. Short interest can proxy for short selling activities. Monthly short interest is shares sold short divided by shares outstanding. Short interest is averaged at 4.55%. Finally, we calculate arbitrage risk, which is mean squared error of residuals (RMSEs) from Carhart's four factor model and has mean values of 2.36%. To coincide with quarterly index holding, all the variables are averaged within quarters. To exclude extreme values, we winsorize the variables at 1^{st} and 99^{th} percentiles. Variable definitions and sources are further documented in Appendix.

Table 1 presents summary statistics from 2002 to 2016 in Panel A and average quarterly correlation in Panel B. On average, index ownership has the same sign and similar magnitude correlation with liquidity and price efficiency as institutional ownership does. Greater passive index holding is related with more liquid and more efficient prices. It also positively relates with short selling activities and negatively with arbitrage risk.

3.3.Empirical Results

In this section, we examine how indexing affects underlying stocks' market quality. We use cross-sectional firm level regressions controlling quarter fixed effects and industry fixed effects, and clustering standard errors on firm level. To improve our ability to identify the consequences of increased indexing in a clean setting, we focus on the effects of lagged changes of passive ownership. In our later robust tests, we also add the impacts of lagged level of passive ownership and the impacts of lagged changes of passive ownership on changes of market quality.

We identify two central dimensions of a firm's information environment: (1) stock liquidity, since lower transaction costs would encourage more trading and arbitrage activities (2) price efficiency, that is the extent or the speed to which stock prices reflect information. In all the regressions, we first employ change of passive ownership without or with change of institutional ownership to make sure that the effect of passive trading is not confounded by the effect of institutional trading. And then we examine two parts related with passive investing: index flows and index rebalancing. Control variables include log of firm size, book-to-market ratio, standard deviation of returns, stock turnover, log of prices, log of number of analysts and lagged period's dependent variable. To mitigate causality issues, all independent variables are lagged by one period. Regression model is as followed:

$$Liquidity\ (Price\ efficiency)_{i,t} = \beta_0 + \beta_1 \Delta Index\%_{i,t-1} + \sum_k \beta_k\, control_{i,t-1} + \mu_i + \eta_t + \varepsilon_{i,t} \quad (9)$$

3.3.1. Passive investing and Stock liquidity

Table 2 presents cross-sectional regression results of bid-ask Spread (column (1) to (4)) and Amihud ratio (column (5) to (8)) on passive investing and controls. In column (1) and (5), we can observe that change of passive ownership is related with decreased bid-ask spread and Amihud ratio in the next quarter, though the impact on Amihud is only statistically significant at 10% confidence interval. Next in column (2) and (6), we additionally control change of institutional ownership, and it doesn't subsume the effects of passive trading. Index trading and institutional trading significantly decrease bid-ask spread in column (2). In specific, a one-standard-deviation increase in index trading (0.0147) is associated with a 0.017% (0.0117*0.0147) decrease in $spread$, which corresponds to1.42% of its standard deviation (0.012). In contrast, a one-standard-deviation change of institutional trading (0.052) only amounts to 0.35% of $spread$'s standard deviation (0.052*0.000812/0.012). So compared with institutional trading effects, which are known to increase stocks' liquidity, index investing has greater economically significant impacts.

But both institutional trading and index trading have no significant coefficients on $Amihud$ in column (6).

Next in column (3) and column (7), we use $stock\_flow$ to proxy for index trading related with fund flows on stock level. The coefficients are significant for both $Spread$ and $Amihud$ under 1% confidence intervals. In specific, with a one-standard-deviation increase in $stock\_flow$ (0.0574), underlying $spread$ decreases by 0.012% (0.00213*0.0574) and underlying $Amihud$ decreases by 6.06% (0.0574*1.057). These reductions correspond to 1.02% in $Spread$'s standard deviation (0.012) and 1.35% in $Amihud$'s standard deviation (4.492). In column (4) and column (8), we divide total passive investing into index flows ($flow\_related$) and index rebalancing ($flow\_residual$), and $flow\_related$ has greater statistical and economical significance on stock liquidity. A one-standard-deviation increase in $flow\_related$ (0.0096) reduces $spread$ by 0.24%, while $flow\_residual$ (0.0112) reduces it by 0.012%. Only $flow\_related$ has significant negative impacts on $Amihud$, while $flow\_residual$ loses statistical significance. Control variables are mostly significant with the expected signs. Spread or Amihud is negatively related to firm size, turnover, stock prices and analysts following, but they have inconsistent signs for return volatility. Generally, our result indicates that larger firms with greater trading activities and more analysts following are more liquid.

Therefore, we provide evidence that greater passive investing decreases transaction costs for underlying stocks, and thus increases liquidity. In passive investing, compared with index rebalancing, index fund flows have larger impacts and more significant influences on component stocks' liquidity. We will continue to test whether passive trading contains information or just transmits liquidity shocks to underlying stocks in the next section.

3.3.2. Passive investing and Price efficiency

Table 3, Table 4, and Table 5 use three measures separately to test how passive investing impacts underlying price efficiency. We continue to separate the tests into four settings: (1) index trading only (2) index trading and institutional trading (3) index investing about fund flows and (4) index investing components including fund flows and fund rebalancing.

In table 3, we first use R^2 measured price delay $Delay\_r$ from column (1) to column (4) and coefficient measured price delay $Delay\_coef$ from column (5) to column (8) separately. We find that passive investing decreases price delay significantly, regardless of which setting we use. In column (1) and column (5), change of index ownership significantly decreases price delay under 1% confidence interval. In column (2) and column (6), controlling institutional trading doesn't alter the observed relation. In specific, in column (2), a one-standard-deviation increase in index trading (0.0147) is associated with decreased $Delay\_r$ by 0.66% (0.451*0.0147). In contrast, institutional trading (0.052) decreases it by 0.16% (0.052*0.0315) only. Similarly, in column (5), index trading decreases $Delay\_coef$ by 1.12% (0.761*0.0147), while institutional trading decreases it by only 0.22% (0.0431*0.052). It suggests that compared with institutional trading, index trading has greater impacts on next quarter price efficiency. Then column (3) (4) and column (7) (8) show that both index flow and index rebalancing decrease stock price delay, but index flows have larger impacts. For example, in column (4), a one-standard-deviation change of $flow\_related$ decreases price delay by 0.24% (0.25*0.0096) while $flow\_residual$ decreases it by 0.01% (0.0112*0.0114). This result supports the view that greater index trading contains market information and increases the speed of market information incorporated into underlying stock prices. Other variables are consistent with the literature: delay decreases with firm size, stock turnover, return volatility, stock prices and analysts following, suggesting that larger and more liquid firms with higher prices and more analysts following tend to have more efficient prices.

In Table 4, we use variance ratios to capture price inefficiency. Consistent with the literature, institutional trading significantly decreases variance ratio and thus increases price efficiency in the next quarter from column (2) to column (4). But when we use change of index ownership in column (1) and column (2), greater index trading is related with higher variance ratio and lower price efficiency, which seems at odds with our above conclusions. While, in column (3), stock level ownership weighted flows significantly decrease variance ratio under 5% confidence interval. As a one-standard-deviation increase in $stock\_flow$ (0.0574), underlying variance ratio decreases by 0.14% (0.0249*0.0574); In contrast, institutional trading (0.0520) decreases variance ratio by 0.23% (0.0520*0.0450). The results in column (4) verify again that $flow\_related$ rather than $flow\_residual$ significantly decreases variance ratio. Investors with fundamental information are likely to trade indexes first, therefore fund flows contain information and would increase underlying stocks' price efficiency through arbitrageurs' activities. Control variables are mostly consistent with the expected signs: variance ratio decreases with firm size, shares turnover and analyst following.

In Table 5, we are attempting to explain impacts of passive investing on the relative amounts of systematic information into stock prices. We control lagged beta as a control in the regression for a firm's systematic risk. In column (1) and (2), results show that greater passive investing significantly increases underlying return synchronicity. With a one-standard-deviation change in $\Delta index\%$, return synchronicity increases by around 4.5% (3.097*0.0147; 3.080*0.0147). In column (3) and (4), both index fund flows and index rebalancing contribute to greater return co-movement. For example, in column (4), $flow\_related$ increases return co-movement by around 0.85 (88.5*0.0096) and $flow\_residual$ increases it by around 0.03 (2.974*0.0112). But in our regression, institutional trading has relatively weak role on underlying stocks' return co-movement, since institutions have discretion about when and what to trade. When we use market and industry returns to construct synchronicity $Sync\_ind$ from column (5) to (8), results are similar. Since both fund flows and fund rebalancing would provide liquidity, and stocks in the same baskets are going to be exposed to the same liquidity shocks, their covariance increases a result. Increased return co-movement suggests that increased price efficiency is driven by systematic information.

Taken together, we provide evidence consistent with benefits to the underlying stocks in the form of better liquidity and better price efficiency. We consider two reasons attributing passive investing: index fund flows and index rebalancing, and test their relative impacts. In general, passive investing associated with fund flows would attract both uninformed liquidity investors and informed investors who impound index-related information into index prices. As such, it is conceivable that index prices reflect an aggregate version of news before it is incorporated into prices of underlying securities. Arbitrageurs who are active in index products help both prices and NAVs to adjust and allow systematic information from passive funds to their underlying securities, causing a closer link between fundamentals and stock prices. In addition, passive investing associated with fund rebalancing would provide liquidity and increase the proportion of available lendable shares at the same time, especially when a stock experiences index addition while doesn't not be excluded from other indexes. Stock liquidity and price efficiency will increase since greater indexing reduces the costs of expressing negative views through short selling activities. We will continue to test the related hypotheses in more details in the following section.

3.4. Channel testing

3.4.1. Arbitrage channel

Open-end mutual funds have daily creation/redemption mechanism that links excess demand or supply from investors, who will at times arbitrarily pick indexes or underlying securities to establish their factor positions and get profits from price discrepancies. Until now, we have showed that greater passive investing, especially fund flows, has significantly increased stock price efficiency. Next, to test arbitrage channel, we employ the environment where greater fund inflows occur. We also differentiate index mutual funds from ETFs, which the latter allows intraday continuous trading.

First, we measure aggregate fund flow at each quarter t as:

$$Agg\_flow_t = \frac{\sum_{j=1}^{J}(TNA_{j,t} - TNA_{j,t-1} * (1 + r_q))}{\sum_{j=1}^{J} TNA_{j,t-1}} \quad (10)$$

Akbas, et al. (2015) use aggregate fund flows to test whether the flows are 'dumb money' which impedes arbitrage function (price pressure effect) or 'smart money' which acts as arbitrage capital and corrects mispricing (information effect). We also exploit this measure, split index investing into high aggregate fund inflow periods and low aggregate fund flow periods. We predict that greater effects should concentrate around high aggregate flow periods if investor flows contain information. In the regression, we do not include level variables in this specification because they are collinear with other include variables and quarter fixed effects.

In table 6, we first investigate the effects of passive investing on idiosyncratic risk. To save spaces, the controls are not reported. In Panel A column (1), change of index ownership doesn't affect arbitrage risk significantly. While in column (2), $\Delta index\%$ interacted with high aggregate fund flow dummy $High\ agg\_flow$ decreases arbitrage risk under 5% confidence interval. In column (3), stock flows significantly decrease underlying stocks' arbitrage risk, which is mostly from high aggregate fund flow periods identified in column (4). In contrast, during low aggregate fund flow periods, both interaction terms from $\Delta index\%$ in column (2) and from $stock\_flow$ in column (4) don't have significant effects. Pontiff (2006) argues that idiosyncratic risk is the largest cost in arbitrage activities, which would limit arbitrageurs' ability to trade on mispricing. Our results suggest that greater passive fund inflows will significantly decrease idiosyncratic risk.

Next in Panel B, we test the relation of how passive investing increases underlying price efficiency under different environments. To save spaces, we only use $delay$ to measure price efficiency from now and don't report control variables. Column (1) and (3) show that change of passive ownership decreases price delay under both high aggregate fund flow periods and low aggregate fund flow periods, but the coefficients of $\Delta index\%_{i,t-1} * High\ agg\_flow$ are larger in magnitude. Column (2) and (4) employ stock-level fund flows as proxy for index trading, and interact with aggregate fund flow dummies respectively. Results show that only during $High\ agg\_flow$, $stock\_flow$ decreases price delay significantly. In contrast, under $low\ agg\_flow$, $stock\_flow$ instead increases price delay.

Second, we select ETFs from total passive funds sample, where ETFs have a share code of '73' in the CRSP or ETF flag 'F'. Compared with index mutual funds, which use the nightly settlement of mutual fund NAVs, ETFs allow intraday trading and are subject to more intense trading, therefore more arbitrages are done by market makers during the day (Ben-David, Franzoni, and Moussawi, 2018). Accordingly, we construct stock level ETF flows $stock\_\ flow\_ETF$ and stock level Index mutual fund flows $stock\_flow\_\ index\ mutual\ funds$.

In Table 7, we follow Ben David, et al. (2018) and test how the ETF holding and index mutual fund holding affects underlying volatility. Consistent with Ben David, et al. (2018), Column (1) shows that passive investing increases underlying volatility, but which is totally driven by ETFs rather than index mutual funds, as shown in Column (2). Similarly, flow induced trading increases underlying volatility in column (3), also driven by ETFs. Panel B shows that the increased volatility will cause increased variance ratio and decreased price efficiency. Therefore, we verify that index mutual funds increase underlying market quality, which will be partially reversed by ETFs' effects.

3.4.2. Short selling channel

Evans et al. (2014) find that passive investors, such as index funds, are more likely to lend securities. Compared with active mutual funds, which prefer to retain stocks that matter to their trading strategy and portfolio performance, passive funds face no adverse selection from providing lending supply (Prado, Saffi and Sturgess, 2016). Therefore, it is conceivable that passive investing affects short selling demands. Consistent with the literature (Boehmer et al. 2008), the relation would be concentrated around negative news event. Greater passive investing affects the speed of negative information flows by relaxing short selling constraints.

Table 8 reports the results about short selling channel, including the same stock controls and not reported. Panel A shows that both index ownership level and index ownership change will increase short selling activities in the next quarter. In specific, in column (1), greater $index\%$ increases short interest by 0.54% (0.139*0.039), while institution ownership increases short interest by 0.8% (0.0284*0.283). Next, column (2) shows that index trading, represented by $\Delta index\%$, increases changes in demand for borrowing stocks. Column (3) and column (4) test two parts of index trading separately. Index trading caused by index rebalancing rather than by fund flows increases short selling demands. For example, in column (4), $flow\_residual$ significantly increases change of short interest by 0.067% (0.0594*0.0112), while $flow\_related$ decreases it. Therefore, if a firm is added to an index while not being dropped from another, the number of shares available for shorting goes up. Increased short selling demands are more related with the shares in passive funds which are available for shorting (Campello and Saffi, 2015).

Next, to test short selling channel, we conduct tests about how passive investing affects the speed of information flows by relaxing short selling constraints. We interact $\Delta index\%$ with short interest and include lagged period's short interest and lagged period's change of passive ownership.

Since we expect that greater short selling demands occur around negative news event, we modify the above regression (5) to isolate negative market returns (Boehmer and Wu, 2013).

$$r_{i,t} = \alpha_i + \beta_i^0 R_{m,t}^- + \sum_{n=1}^{5} \beta_i^n R_{m,t-n}^- + \varepsilon_{i,t} \tag{11}$$

where $R_{m,t}^-$ equals the daily market return when it is negative. We then use R^2 and coefficients to construct $delay\_r\_down$ and $delay\_coef\_down$. To compare the effects, we also isolate positive market returns and calculate $delay\_r\_up$ and $delay\_coef\_up$.

Panel B shows that in all the specifications, short interest is negatively related with price delay and thus increases price efficiency. But only when we use negative market returns to construct price delay in column (3) and (4), interaction terms $\Delta index\%_{i,t-1} * short\ interest_{i,t-1}$are negatively significant. It means that given negative news in the market, passive investing will increase the speed of market information incorporated into stock prices by allowing greater short selling activities. Especially, in panel C, after we spilt sample equally each quarter by beginning level of arbitrage risk, the interaction terms are only significant for the stocks with high arbitrage risk. Hence, this effect especially benefits the hard to borrow stocks.

Finally, in this section, we also consider the difference in underlying markets. In the absence of index funds, investors who trade relatively illiquid assets directly will incur potential large transaction costs. Since small and less liquid have less public available information, information asymmetry among market participants may be substantial. Therefore, limits to arbitrage and constraints to short selling should be greater. But in the presence of indexers, investors can trade a basket of securities without trading the underlying stocks, which allows more arbitrage activities on these stocks and thus makes information processed faster. Moreover, indexing increases the visibility and lendable shares for the stocks which were difficult to borrow under market, so we will observe greater trading volumes attributed to these stocks. Therefore, we assume that greater passive investing will increase stocks' liquidity and increase stock price efficiency especially for the stocks which were hard to trade in the market before.

In Table 9, we assess how the relation between passive investing and market quality varies with stock size (Panel A) and stock turnover (Panel B). For each measure, we first sort total sample each quarter into terciles, and run regressions separately within these three groups. To save spaces, we only report coefficients on passive investing. In the last column, we test the significance for coefficients' differences between the largest group and the smallest group using $x^2$and p values.

In Panel A, we can observe that change of passive ownership increases next quarter's bid-ask spread and price delay in a larger magnitude for small size stocks than for large size stocks. The differences in the last column are significant. The only exception is when we regress $Amihud$ on $\Delta index\%$ for different groups of stocks and the difference in the last column is not significant either. Next when we use stock level ownership weighted fund flows to proxy for passive trading, it only increases stock liquidity and price efficiency for the small size stocks. In contrast, price delay will even increase and price efficiency decrease for the large size stocks. Therefore, passive investing's positive effect on underlying market quality is concentrated on the small stocks. Similar conclusions apply to Panel B when we divide sample by liquidity proxy. But the differences are only significant when testing liquidity. When we test how price delay varies for stocks with different turnovers, differences in the last column are not significant.

3.5.Robust tests

3.5.1 Propensity score matching

Stocks with high passive investing and stocks with low passive investing may differ along some other dimensions correlated with stock liquidity and price efficiency. To ensure that our results are not driven by such a bias, we employ a propensity score matched sample specification.

For passive investing, including change of passive ownership and stock level ownership weighted fund flows, we identify a treatment group and a control group that are similar along each given characteristic and test the differences of stock liquidity and price efficiency.

We match on the propensity of a stock having high index trading, where high index trading is classified as a dummy variable which equals to one if stocks are in the top quartile $\Delta index\%$ (Panel A) or $stock\_flow$ (Panel B) each quarter. Then in the first stage, we compute propensity scores using a logit regression. We regress high index trading dummy on the same controls as the above regression setting, including stock size, book to market, prices, turnover, return volatility, bid-ask spread, lagged period's index holding, lagged period's institutional holding, and high orders of these covariates or their interactions. Consistent with the above setting, we also control industry and quarter fixed effects, and cluster standard errors across individual stocks. Next, the propensity scores are calculated from a 1:1 matching without replacement and with a 0.01 caliper.

Table 10 Panel A reports differences of stock liquidity and price efficiency between treatment group and control group. First, when we use $\Delta index\%$, there are not significant differences for liquidity measures such as $spread$ in column (1) and $Amihud$ in column (2). But when we measure price delay in column (3) and column (4), the differences are significant under 1% confidence interval. Second, when we use $stock\_flow$, $spread$ of two matched groups are significantly different under 5% confidence interval and price delay differences are significant under 1% confidence intervals. These results suggest that stocks with high passive investing would have higher price efficiency significantly.

3.5.2 Index trading and Index holding

We conduct tests about how index holding ($index\%$) impacts price efficiency and how index trading ($\Delta index\%$) impacts change of price efficiency. Table 10 Panel B reports the results. In column (1), we regress price delay on lagged period's index ownership, controlling lagged period's institutional ownership and stock controls. We find that $index\%$ significantly decreases price delay. A one-standard-deviation increase in $index\%$ (0.039) decreases next quarter price delay by 3.93% (1.009*0.039). In contrast, institutional holdings (0.283) decrease price delay by 1.54% (0.0543*0.283). So compared with large institutions, passive indexes will statistically and economically improve price efficiency of the component stocks. In column (2) and column (3), we regress change of price delay on lagged period's passive trading, institutional trading and changes of stock controls. The results show that both $\Delta index\%$ and $stock\_flow$ significantly decrease change of price delays in the next quarter, which means that greater passive investing leads to an increase in greater underlying price efficiency.

3.5.3 Other regression specifications

Table 10 Panel C compares panel regression with Fama-Macbeth regression. First, column (1) reports the results using panel regression controlling firm fixed effects and quarter fixed effects and clustering standard errors on individual firms. Next in column (2), we use quarterly Fama-Macbeth regression. Both of the two regressions report results that passive trading significantly decreases price delay and increases price efficiency. In fact, institutional trading is not significant in either of these two models.

3.5.4 Reverse causality

Large indexes will mechanically hold efficiently priced stocks. Therefore, we want to show whether our results are impacted by reverse causality. We employ Granger causality tests and regress time-series changes in price efficiency on lagged time-series changes in passive ownership for each stock and also control lagged changes of stock controls. Table 10 Panel D shows that greater passive trading is positively associated with a greater improvement in price efficiency in column (1). To examine reverse causality, we regress changes in passive ownership on lagged changes in price efficiency and control the same lagged changes of stock controls. Column (2) shows that the average coefficient on lagged price delay is not significantly related with changes in passive trading. Therefore, our results are not easily explainable using a reverse causality story.

3.5.5 Time periods

Finally, we want to see whether our results only focus on some certain time periods. We first divide sample periods into before 2008 and after 2008, since passive investing increases intensely after Financial Crisis. In table 10 panel E we can observe that there are not significantly differences between two periods in column (1) and (2). The results are similar that greater passive investing is associated with increased price efficiency.

Next, we divide sample periods by CBOE volatility index (VIX). We select VIX median in our sample periods, which is 18%, to divide sample into two sub-periods. In column (3) and column (4), our results are not affected by total market volatility.

3.6 Conclusion

The rapid growth of the indexing industry, including index mutual funds and (index) ETFs, has attracted a lot of attention from academics, practitioners, and regulators over the past two decades. How indexing impacts the underlying market quality has become a major concern.

In this paper, we first find that indexing increases underlying stocks' liquidity and price efficiency. The increase in price efficiency is mostly from systematic information rather than from firm specific information, and mostly concentrated in stocks that are difficult to trade, such as smaller stocks and less liquid stocks. By analyzing reasons contributing to passive trading, we construct stock-level proxies for index flows and index rebalancing. On one hand, greater index flows create price discrepancies between indexes and constituent stocks, which encourages arbitrageurs to transmit systematic information into underlying prices. On the other hand, stock addition into indexes increases the available lendable shares and allows short sellers to express negative news and correct overvaluation. To conclude, indexing attracts both liquidity investors who attempt to achieve diversification and informed investors who trade with macro-based information, which benefits to the underlying stocks in the form of better liquidity and better price efficiency. We emphasize arbitrage and short selling as the driver of liquidity and price efficiency.

References

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

Akbas, F., Boehmer, E., Erturk, B., & Sorescu, S. (2017). Short interest, returns, and unfavorable fundamental information. Financial Management, 46(2), 455-486.

Appel, I. R., Gormley, T. A., & Keim, D. B. (2016). Passive investors, not passive owners. Journal of Financial Economics, 121(1), 111-141.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. Journal of financial markets, 5(1), 31-56.

Ben-David, I., Franzoni, F. and Moussawi, R., 2018. Do ETFs increase volatility?. *The Journal of Finance*, *73*(6), pp.2471-2535.

Ben-David, I., Franzoni, F., & Moussawi, R. (2016). Exchange Traded Funds (ETFs) (No. w22829). National Bureau of Economic Research.

Bessembinder, H. (2015). Predictable ETF Order Flows and Market Quality. The Journal of Trading, 10(4), 17-23.

Bhattacharya, A. and O'Hara, M., 2018. Can ETFs increase market fragility? Effect of information linkages in ETF markets. *Effect of Information Linkages in ETF Markets (April 17, 2018)*.

Blocher, J., & Whaley, R. E. (2016). Two-Sided Markets in Asset Management: Exchange-Traded Funds and Securities Lending.

Boehmer, B., & Boehmer, E. (2003). Trading your neighbor’s ETFs: Competition or fragmentation?. Journal of Banking & Finance, 27(9), 1667-1703.

Boehmer, E., & Kelley, E. K. (2009). Institutional investors and the informational efficiency of prices. Review of Financial Studies, 22(9), 3563-3594.

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

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

Broman, M. S., & Shum, P. M. (2016). Short-Term Trading and Liquidity Clienteles: Evidence from Exchange-Traded Funds.

Brown, D.C., Davies, S. and Ringgenberg, M., 2018. ETF Flows, Non-Fundamental Demand, and Return Predictability. *Available at SSRN 2872414*.

Campello, M., & Saffi, P. A. (2015). The rise of the equity lending market: Implications for corporate financial policies.

Chen, H., Noronha, G., & Singal, V. (2004). The price response to S&P 500 index additions and deletions: Evidence of asymmetry and a new explanation. The Journal of Finance, 59(4), 1901-1930.

Chordia, T., Roll, R., & Subrahmanyam, A. (2008). Liquidity and market efficiency. Journal of Financial Economics, 87(2), 249-268.

Cong, L. W., & Xu, D. (2016). Rise of factor investing: Asset prices, informational efficiency, and security design.

Coval, J., & Stafford, E. (2007). Asset fire sales (and purchases) in equity markets. Journal of Financial Economics, 86(2), 479-512.

Crane, A. D., & Crotty, K. (2016). Passive versus Active Fund Performance: Do Index Funds Have Skill?.

Crawford, S. S., Roulstone, D. T., & So, E. C. (2012). Analyst initiations of coverage and stock return synchronicity. The Accounting Review, 87(5), 1527-1553.

Da, Z. and Shive, S., 2018. Exchange traded funds and asset return correlations. *European Financial Management*, *24*(1), pp.136-168.

Evans, R., Ferreira, M. A., & Porras Prado, M. (2017). Fund Performance and Equity Lending: Why Lend What You Can Sell?. Review of Finance, 21(3), 1093-1121.

Fremault, A. (1991). Stock index futures and index arbitrage in a rational expectations model. Journal of Business, 523-547.

French, K. R. (2008). Presidential address: The cost of active investing. The Journal of Finance, 63(4), 1537-1573.

Glosten, L. R., Nallareddy, S., & Zou, Y. (2016). ETF trading and informational efficiency of underlying securities.

Gorton, G., & Pennacchi, G. (1991). Security baskets and index-linked securities (No. w3711). National Bureau of Economic Research.

Hamm, S. J. (2014). The effect of ETFs on stock liquidity.

Hou, K., & Moskowitz, T. J. (2005). Market frictions, price delay, and the cross-section of expected returns. Review of Financial Studies, 18(3), 981-1020.

Israeli, D., Lee, C. M., & Sridharan, S. A. (2016). Is there a dark side to exchange traded funds? An information perspective. Review of Accounting Studies, 1-36.

Kumar, P., & Seppi, D. J. (1994). Information and index arbitrage. Journal of Business, 481-509.

Li, F.W. and Zhu, Q., 2018. Short selling etfs. *Available at SSRN 2836518*.

Lo, A. W., & MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. The review of financial studies, 1(1), 41-66.

Madhavan, A., & Sobczyk, A. (2015). Price dynamics and liquidity of exchange-traded funds.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. The Journal of Finance, 42(3), 483-510.

Nagel, S., 2005. Short sales, institutional investors and the cross-section of stock returns. *Journal of financial economics*, *78*(2), pp.277-309.

Piotroski, J. D., & Roulstone, D. T. (2004). The influence of analysts, institutional investors, and insiders on the incorporation of market, industry, and firm-specific information into stock prices. The Accounting Review, 79(4), 1119-1151.

Porras Prado, M., Saffi, P. A., & Sturgess, J. (2016). Ownership Structure, Limits to Arbitrage, and Stock Returns: Evidence from Equity Lending Markets. The Review of Financial Studies, 29(12), 3211-3244.

Qin, N., & Singal, V. (2015). Indexing and stock price efficiency. Financial Management, 44(4), 875-904.

Saffi, P. A., & Sigurdsson, K. (2010). Price efficiency and short selling. The Review of Financial Studies, 24(3), 821-852.

Sirri, E. R., & Tufano, P. (1998). Costly search and mutual fund flows. Journal of Finance, 53(5), 1589-1622.

Stambaugh, R. F. (2014). Presidential address: Investment noise and trends. The Journal of Finance, 69(4), 1415-1453.

Subrahmanyam, A. (1991). A theory of trading in stock index futures. Review of Financial Studies, 4(1), 17-51.

Sullivan, R. N., & Xiong, J. X. (2012). How index trading increases market vulnerability. Financial Analysts Journal, 68(2), 70-84.

Wurgler, J. (2010). On the economic consequences of index-linked investing (No. w16376). National Bureau of Economic Research.

Table 1: Summary Statistics

Table 1 presents summary statistics of U.S. common stock traded on NYSE, AMEX, and Nasdaq from 2002.q1 to 2016.q4. All the variables are winsorized at 1st and 99th percentiles. Variable definitions are provided in Appendix A. We require that stocks in the sample have available data about prices and shares outstanding, have prices between $1 and $1000, and have positive index holdings and institutional holdings. In addition, we require that there are at least 50 observations per firm-quarter to calculate beta, return synchronicity and price delay.

Panel A: Summary Statistics

Variable	Mean	Median	Std deviation
Ownership			
Index%	0.0488	0.0369	0.0438
Num_index	39.2280	32.0000	33.1475
Inst%	0.5346	0.5795	0.275
Num_inst	129.9421	87.0000	152.9612
Stock_flow	0.0264	0.0199	0.0497
Price Efficiency			
Delay_r	0.3819	0.2747	0.3084
Delay_coef	1.6679	1.5792	0.6479
Variance ratio	0.3376	0.3008	0.2363
Co-movement			
Beta	1.0156	0.9924	0.6937
Sync_mkt	-2.0768	-1.5441	2.0734
Sync_ind	-1.3851	-1.1231	1.5465
Liquidity			
Spread	0.0071	0.0021	0.0120
Amihud	1.0554	0.0067	4.8687
Control variables			
Size	3520.5600	496.1769	10160.8200
Btm	0.7068	0.5555	0.6016
Price	24.6299	17.6500	23.8246
Turnover	0.0080	0.0057	0.0080
Std(ret)	0.0285	0.0241	0.0170
Arbi_risk	0.0236	0.0214	0.0061
Short interest	0.0455	0.0273	0.0557
Analyst	7.0946	5.0000	7.2353

Panel B: Average quarterly correlation of ownership with stock liquidity and price efficiency

	Delay_r	Delay_coef	Variance_	Beta	Sync_mkt	Sync_ind	Spread	Amihud	Short int
Index%	-0.327	-0.281	-0.072	0.275	0.361	0.375	-0.381	-0.334	0.384
Inst%	-0.374	-0.322	-0.12	0.309	0.392	0.386	-0.564	-0.549	0.451

Table 2: Passive Investing and Stock Liquidity

Table2 reports stock-level panel regression of quarterly averaged *bid-ask Spread* and *Amihud ratio* on *Passive investing* and control variables for U.S. common stocks. Passive investing includes *change of passive ownership* in column (1) (2) and column (5) (6), *ownership weighted fund flows* in column (3) and column (7), *index trading components* in column (4) and column (8). All the independent variables are lagged by one period. Variable definitions are in Appendix A. Panel regression controls quarter fixed effects and industry fixed effects, and standard errors are clustered at stock level. T statistics are provided in parentheses, and *, **, *** represent significance at the 10%, 5%, and 1% confidence intervals.

	$Y = Spread_t$				$Y = Amihud_t$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta index\%_{t-1}$	-0.0119***	-0.0117***			-0.990*	-0.956		
	(-13.56)	(-13.32)			(-1.66)	(-1.60)		
$stock\_flow_{t-1}$			-0.00213***				-1.057***	
			(-9.89)				(-6.64)	
$flow\_related_{t-1}$				-0.250***				-124.1***
				(-9.92)				(-6.64)
$flow\_residual_{t-1}$				-0.0114***				-0.841
				(-13.12)				(-1.41)
$\Delta inst\%_{t-1}$		-0.000812***	-0.000943***	-0.000804***		-0.138	-0.157	-0.146
		(-5.00)	(-5.78)	(-4.93)		(-1.24)	(-1.41)	(-1.31)
$size_{t-1}$(log)	-0.000359***	-0.000359***	-0.000359***	-0.000360***	-0.173***	-0.173***	-0.173***	-0.173***
	(-24.79)	(-24.83)	(-24.82)	(-24.92)	(-14.66)	(-14.67)	(-14.63)	(-14.64)
btm_{t-1}	-0.0000431***	-0.0000434***	-0.0000411***	-0.0000416***	-0.0103***	-0.0104***	-0.00960***	-0.00963***
	(-13.95)	(-14.02)	(-13.32)	(-13.48)	(-3.90)	(-3.91)	(-3.59)	(-3.60)
$std(ret)_{t-1}$	-0.0300***	-0.0301***	-0.0296***	-0.0299***	7.374***	7.362***	7.451***	7.433***
	(-17.51)	(-17.55)	(-17.28)	(-17.43)	(5.18)	(5.17)	(5.22)	(5.21)
$turnover_{t-1}$	-0.00529***	-0.00514***	-0.00558***	-0.00519***	-20.89***	-20.86***	-20.83***	-20.81***
	(-3.27)	(-3.17)	(-3.43)	(-3.20)	(-13.58)	(-13.55)	(-13.55)	(-13.54)
$price_{t-1}$(log)	-0.000201***	-0.000200***	-0.000198***	-0.000198***	-0.0537***	-0.0534***	-0.0521***	-0.0521***
	(-9.23)	(-9.15)	(-9.05)	(-9.05)	(-2.98)	(-2.97)	(-2.89)	(-2.88)
$analyst_{t-1}$(log)	-0.000151***	-0.000153***	-0.000158***	-0.000153***	-0.0912***	-0.0914***	-0.0915***	-0.0911***
	(-8.61)	(-8.71)	(-8.92)	(-8.68)	(-6.28)	(-6.29)	(-6.28)	(-6.26)
dv_{t-1}	0.876***	0.876***	0.875***	0.875***	0.640***	0.640***	0.640***	0.640***
	(290.88)	(290.82)	(289.68)	(289.99)	(63.40)	(63.40)	(63.26)	(63.26)
Intercept	0.00460***	0.00460***	0.00476***	0.00192***	1.903***	1.905***	1.914***	2.930***
	(7.20)	(7.20)	(7.44)	(2.88)	(7.90)	(7.92)	(7.96)	(9.99)

Ind FE	YES	YES	YES	YES	YES	YES	YES	YES
Qr FE	YES	YES	YES	YES	YES	YES	YES	YES
N	168746	168746	168431	168431	168773	168773	168458	168458
R^2	0.866	0.866	0.866	0.866	0.501	0.501	0.502	0.502

Table 3: Passive Investing and Price Efficiency (Price Delay)

Table3 reports stock-level panel regression of *R-square measured Price Delay* and *Coefficient measured Price Delay* on *Passive investing* and control variables for U.S. common stocks. Passive investing includes *change of passive ownership* in column (1) (2) and column (5) (6), *ownership weighted fund flows* in column (3) and column (7), *index trading components* in column (4) and column (8). All the independent variables are lagged by one period. Variable definitions are in Appendix A. Panel regression controls quarter fixed effects and industry fixed effects, and standard errors are clustered at stock level. T statistics are provided in parentheses, and *, **, *** represent significance at the 10%, 5%, and 1% confidence intervals.

	$Y = Delay\_r_t$				$Y = Delay\_coef_t$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta index\%_{t-1}$	-0.458***	-0.451***			-0.772***	-0.761***		
	(-9.62)	(-9.45)			(-7.03)	(-6.92)		
$stock\_flow_{t-1}$			-0.0827***				-0.138***	
			(-25.56)				(-18.96)	
$flow\_related_{t-}$				-8.556***				-14.30***
				(-25.60)				(-18.97)
$flow\_residual_{,}$				-0.262***				-0.445***
				(-5.40)				(-3.99)
$\Delta inst\%_{t-1}$		-0.0315***	-0.0433***	-0.0400***		-0.0431*	-0.0644***	-0.0588**
		(-2.89)	(-3.97)	(-3.66)		(-1.75)	(-2.62)	(-2.38)
$size_{t-1}$(log)	-0.0438***	-0.0438***	-0.0441***	-0.0441***	-0.0976***	-0.0976***	-0.0980***	-0.0980***
	(-36.90)	(-36.93)	(-37.96)	(-37.98)	(-37.47)	(-37.49)	(-38.19)	(-38.20)
btm_{t-1}	-0.00404***	-0.00405***	-0.00444***	-0.00444***	-0.00908***	-0.00909***	-0.00970***	-0.00970***
	(-12.66)	(-12.70)	(-13.84)	(-13.83)	(-12.05)	(-12.07)	(-12.82)	(-12.81)
$std(ret)_{t-1}$	-0.887***	-0.889***	-0.922***	-0.926***	-1.550***	-1.553***	-1.611***	-1.619***
	(-11.85)	(-11.88)	(-12.41)	(-12.47)	(-9.04)	(-9.06)	(-9.46)	(-9.50)
$turnover_{t-1}$	-1.139***	-1.133***	-1.016***	-1.012***	-1.871***	-1.862***	-1.650***	-1.642***
	(-8.59)	(-8.54)	(-7.74)	(-7.71)	(-6.18)	(-6.15)	(-5.48)	(-5.46)
$price_{t-1}$(log)	-0.0222***	-0.0221***	-0.0214***	-0.0214***	-0.0445***	-0.0444***	-0.0433***	-0.0433***
	(-13.60)	(-13.56)	(-13.33)	(-13.34)	(-12.16)	(-12.13)	(-11.96)	(-11.96)
$analyst_{t-1}$(log	-0.0175***	-0.0175***	-0.0166***	-0.0165***	-0.0346***	-0.0347***	-0.0330***	-0.0329***
	(-10.74)	(-10.77)	(-10.32)	(-10.28)	(-9.65)	(-9.66)	(-9.28)	(-9.24)
dv_{t-1}	0.374***	0.374***	0.370***	0.371***	0.261***	0.261***	0.259***	0.259***
	(96.12)	(96.12)	(96.31)	(96.35)	(72.47)	(72.47)	(72.51)	(72.53)
Intercept	0.551***	0.551***	0.549***	0.627***	1.927***	1.928***	1.924***	2.054***
	(22.53)	(22.59)	(22.53)	(25.57)	(36.12)	(36.17)	(36.01)	(38.23)

Ind FE	YES	YES	YES	YES	YES	YES	YES	YES
Qtr FE	YES	YES	YES	YES	YES	YES	YES	YES
N	168773	168773	168458	168458	168773	168773	168458	168458
R^2	0.481	0.481	0.483	0.483	0.376	0.376	0.377	0.377

Table 4: Passive Investing and Price Efficiency (Variance Ratio)

Table4 reports stock-level panel regression of *Variance ratio* on *Passive investing* and control variables for U.S. common stocks. Passive investing includes *change of passive ownership* in column (1) (2), *ownership weighted fund flows* in column (3), *index trading components* in column (4). All the independent variables are lagged by one period. Variable definitions are in Appendix A. Panel regression controls quarter fixed effects and industry fixed effects, and standard errors are clustered at stock level. T statistics are provided in parentheses, and *, **, *** represent significance at the 10%, 5%, and 1% confidence intervals.

	$Y = Variance\ ratio_t$			
	(1)	(2)	(3)	(4)
$\Delta index\%_{t-1}$	0.0996**	0.111**		
	(2.02)	(2.24)		
$stock\_flow_{t-1}$			-0.0249**	
			(-2.19)	
$flow\_related_{t-1}$				-2.918**
				(-2.18)
$flow\_residual_{t-1}$				0.114**
				(2.31)
$\Delta inst\%_{t-1}$		-0.0450***	-0.0450***	-0.0464***
		(-4.10)	(-4.10)	(-4.22)
$size_{t-1}$(log)	-0.0125***	-0.0125***	-0.0125***	-0.0125***
	(-16.66)	(-16.69)	(-16.66)	(-16.65)
btm_{t-1}	-0.00159***	-0.00160***	-0.00156***	-0.00156***
	(-6.68)	(-6.74)	(-6.45)	(-6.43)
$std(ret)_{t-1}$	0.292***	0.288***	0.290***	0.292***
	(4.52)	(4.46)	(4.48)	(4.52)
$turnover_{t-1}$	-1.537***	-1.528***	-1.533***	-1.537***
	(-15.62)	(-15.53)	(-15.55)	(-15.58)
$price_{t-1}$(log)	0.00276**	0.00285**	0.00292***	0.00291***
	(2.47)	(2.55)	(2.61)	(2.60)
$analyst_{t-1}$(log)	-0.0121***	-0.0121***	-0.0120***	-0.0121***
	(-11.22)	(-11.29)	(-11.18)	(-11.22)
dv_{t-1}	0.0420***	0.0420***	0.0419***	0.0419***
	(15.03)	(15.04)	(14.97)	(14.98)
Intercept	0.443***	0.443***	0.445***	0.469***
	(23.99)	(24.07)	(24.20)	(21.63)
Ind FE	YES	YES	YES	YES
Qtr FE	YES	YES	YES	YES
N	168773	168773	168458	168458
R^2	0.052	0.052	0.052	0.052

Table 5: Passive Investing and Systematic Price Efficiency

Table5 reports stock-level panel regression of *Return Synchronicity* using *market* returns and *Return Synchronicity* using *market and industry* returns on *Passive investing* and control variables for U.S. common stocks. Passive investing includes *change of passive ownership* in column (1) (2) and column (5) (6), *ownership weighted fund flows* in column (3) and column (7), *index trading components* in column (4) and column (8). All the independent variables are lagged by one period. Variable definitions are in Appendix A. Panel regression controls time fixed effects and industry fixed effects, and standard errors are clustered at stock level. T statistics are provided in parentheses, and *, **, *** represent significance at the 10%, 5%, and 1% confidence intervals.

	$Y = Sync\_mkt_t$				$Y = Sync\_ind_t$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta index\%_{t-1}$	3.097***	3.080***			2.510***	2.512***		
	(10.28)	(10.21)			(12.20)	(12.22)		
$stock\_flow_{t-1}$			0.753***				0.751***	
			(9.43)				(12.27)	
$flow\_related_{t-1}$				88.50***				88.22***
				(9.43)				(12.29)
$flow\_residual_{t-1}$				2.974***				2.382***
				(9.84)				(11.60)
$\Delta inst\%_{t-1}$		0.0717	0.121*	0.0851		-0.00735	0.0254	-0.00343
		(1.04)	(1.76)	(1.24)		(-0.15)	(0.52)	(-0.07)
$size_{t-1}$(log)	0.364***	0.364***	0.363***	0.363***	0.274***	0.274***	0.273***	0.273***
	(41.50)	(41.50)	(41.44)	(41.51)	(36.19)	(36.19)	(36.23)	(36.29)
btm_{t-1}	0.0358***	0.0358***	0.0352***	0.0353***	0.0147***	0.0147***	0.0142***	0.0143***
	(15.11)	(15.13)	(14.81)	(14.86)	(7.32)	(7.31)	(7.07)	(7.12)
$std(ret)_{t-1}$	-15.79***	-15.78***	-15.89***	-15.82***	-9.847***	-9.848***	-9.953***	-9.899***
	(-24.67)	(-24.65)	(-24.76)	(-24.67)	(-19.66)	(-19.66)	(-19.92)	(-19.82)
$turnover_{t-1}$	-2.713***	-2.727***	-2.709***	-2.802***	0.674	0.675	0.652	0.577
	(-2.93)	(-2.95)	(-2.91)	(-3.01)	(0.85)	(0.85)	(0.82)	(0.72)
$price_{t-1}$(log)	0.161***	0.161***	0.160***	0.160***	0.0924***	0.0924***	0.0916***	0.0916***
	(13.33)	(13.31)	(13.25)	(13.25)	(8.88)	(8.88)	(8.81)	(8.81)
$analyst_{t-1}$(log)	0.131***	0.131***	0.132***	0.130***	0.135***	0.135***	0.136***	0.135***
	(10.91)	(10.91)	(10.93)	(10.84)	(13.40)	(13.39)	(13.49)	(13.41)
$beta_{t-1}$	1.050***	1.050***	1.050***	1.050***	0.777***	0.777***	0.778***	0.778***
	(84.91)	(84.91)	(84.82)	(84.86)	(76.72)	(76.73)	(76.84)	(76.90)
Intercept	-4.878***	-4.880***	-4.862***	-5.588***	-3.310***	-3.310***	-3.297***	-4.020***
	(-23.62)	(-23.64)	(-23.56)	(-24.95)	(-17.65)	(-17.65)	(-17.46)	(-20.36)

Ind FE	YES	YES	YES	YES	YES	YES	YES	YES
Qtr FE	YES	YES	YES	YES	YES	YES	YES	YES
N	168773	168773	168458	168458	168773	168773	168458	168458
R^2	0.496	0.496	0.498	0.498	0.537	0.537	0.538	0.538

Table 6: Arbitrage Channel (Arbitrage risk)

Table6 reports stock-level panel regression of *Arbitrage risk* (Panel A), and *Price Efficiency* (Panel B) on *Passive investing, interaction terms of dummy variables* identifying high aggregate flow periods and low aggregate flow periods, and control variables (untabulated) for U.S. common stocks. Passive investing includes *change of index ownership* and *index ownership weighted fund flows*. Control variables are defined as above, but also include last period's arbitrage risk. *Aggregate fund flows* are measured as: $agg\_flow_t = \frac{\sum_{j=1}^{J}(TNA_{j,t} - TNA_{j,t-1} * (1+r_q))}{\sum_{j=1}^{J} TNA_{j,t-1}}$, and are used to divide sample periods into halves. All the independent variables are lagged by one period. Variable definitions are in Appendix A. Panel regression controls time fixed effects and industry fixed effects, and standard errors are clustered at stock level. T statistics are provided in parentheses, and *, **, *** represent significance at the 10%, 5%, and 1% confidence interval.

Panel A: Passive investing and Arbitrage risk

	$\boldsymbol{Y = Arbi\_risk_t}$			
	(1)	(2)	(3)	(4)
$\Delta index\%_{t-1}$	-0.000413			
	(-0.21)			
$\Delta index\%_{t-1}$ $* \, High\ agg\_flow_{t-1}$		-0.00579**		
		(-2.22)		
$\Delta index\%_{t-1}$ $* \, Low\ agg\_flow_{t-1}$		0.00543*		
		(1.90)		
$stock\_flow_{t-1}$			-0.00175***	
			(-3.78)	
$stock\_flow_{t-1}$ $* \, High\ agg\_flow_{t-1}$				-0.00376***
				(-5.62)
$stock\_flow_{t-1}$ $* \, Low\ agg\_flow_{t-1}$				-0.000627
				(-0.98)
$Controls$	Yes	Yes	Yes	Yes
Ind FE	YES	YES	YES	YES
Qtr FE	YES	YES	YES	YES
N	166771	166771	166465	166465
R^2	0.624	0.625	0.624	0.612

Panel B: Passive investing and Price Efficiency brought by Arbitrage activities

	Y= $\boldsymbol{Delay\_coef_t}$			
	(1)	(2)	(3)	(4)
$\Delta index\%_{t-1}$ $* \, High \; agg\_flow_{t-1}$	-0.706***		-1.351***	
	(-10.76)		(-8.79)	
$\Delta index\%_{t-1}$ $* \, Low \; agg\_flow_{t-1}$	-0.290***		-0.422***	
	(-4.54)		(-2.82)	
$stock\_flow_{t-1}$ $* \, High \; agg\_flow_{t-1}$		-0.183***		-0.251***
		(-11.63)		(-7.17)
$stock\_flow_{t-1}$ $* \, Low \; agg\_flow_{t-1}$		0.0496***		0.115***
		(3.00)		(3.05)
$Controls$	Yes	Yes	Yes	Yes
Ind FE	YES	YES	YES	YES
Qtr FE	YES	YES	YES	YES
N	166771	166465	166771	166465
R^2	0.484	0.484	0.389	0.389

Table 7: Arbitrage Channel (ETFs *vs.* Index mutual funds)

Table 7 reports stock-level panel regression of *Daily volatility* (Panel A), and *Price Efficiency* (Panel B) on *Passive investing* and control variables (untabulated) for U.S. common stocks. I divide total passive indexes into index mutual funds and ETFs (CRSP share code “73” or ETF_flag “F”) and construct *ETF ownership* (*ETF flows*) and *index mutual fund ownership (index mutual fund flows),* separately. All the independent variables are lagged by one period. Variable definitions are in Appendix A. Panel regression controls time fixed effects and stock fixed effects, and standard errors are clustered at stock level. T statistics are provided in parentheses, and *, **, *** represent significance at the 10%, 5%, and 1% confidence interval.

Panel A: ETF ownership, Index mutual fund ownership and Daily volatility

	$\boldsymbol{Y = Daily\,Volatility_t}$			
	(1)	(2)	(3)	(4)
$index\%_{t-1}$	0.0185***			
	(10.96)			
$index\ fund\%_{t-1}$		0.00236		
		(0.63)		
$ETF\%_{t-1}$		0.0263***		
		(11.25)		
$stock\_flow_{t-1}$			0.00179***	
			(2.64)	
$stock\_flow\_indexfund_{i,t-}$				-0.00425***
				(-5.75)
$stock\_flow\_ETF_{t-1}$				0.00250***
				(5.06)
$inst\%_{t-1}$	-0.00123***	-0.00126***	-0.000355	-0.000203
	(-3.88)	(-3.97)	(-1.16)	(-0.61)
$Controls$	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Qtr FE	Yes	Yes	Yes	Yes
N	187989	187989	187980	170225
R^2	0.535	0.535	0.535	0.556

Panel B: ETF ownership, Index mutual fund ownership and Price Efficiency

	$Y = Delay\_r_t$		$Y = Variance\ Ratio_t$	
	(1)	(2)	(3)	(4)
$index\ fund\%_{t-1}$	-0.302***		-0.0765	
	(-3.77)		(-1.08)	
$ETF\%_{t-1}$	-0.345***		0.130***	
	(-6.71)		(2.85)	
$stock\_flow\_indexfund_{t-1}$		0.0247		-0.0165
		(1.56)		(-0.97)
$stock\_flow\_ETF_{t-1}$		-0.0114		0.0291***
		(-1.19)		(2.70)
$inst\%_{t-1}$	-0.0826***	-0.0946***	-0.0217***	-0.0216***
	(-11.38)	(-12.64)	(-3.63)	(-3.39)
$Controls$	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Qtr FE	Yes	Yes	Yes	Yes
N	187989	170225	187989	170225
R^2	0.150	0.146	0.017	0.018

Table 8: Short Selling Channel

Table8 reports stock-level panel regression to test *short selling activities* brought by *Passive investing* for U.S. common stocks. Panel A regresses short interest on index holdings in column (1), and regresses changes of short interest on index trading from column (2) to column (4). Panel B regresses price delay on passive investing, interaction terms of short interest and passive investing. The dependent variable in Column (3) and (4) is a modified price delay that uses only down-market returns. The dependent variable in Column (5) and (6) uses only up-market returns. Panel C uses the same setting as panel B, but divides sample equally into triples at the beginning of each quarter by arbitrage risk. All the controls are the same as before and not tabulated. All the independent variables are lagged by one period. Variable definitions are in Appendix A. Panel regression controls time fixed effects and industry fixed effects, and standard errors are clustered at stock level. T statistics are provided in parentheses, and *, **, *** represent significance at the 10%, 5%, and 1% confidence interval.

Panel A: Passive indexes and Short Selling activities

	$Y = Short\ Interest_t$	$Y = \Delta Short\ Interest_t$		
	(1)	(2)	(3)	(4)
$index\%_{t-1}$	0.139***			
	(10.48)			
$inst\%_{t-1}$	0.0284***			
	(14.14)			
$\Delta index\%_{t-1}$		0.0575***		
		(12.59)		
$stock\_flow_{t-1}$			-0.00977***	
			(-11.82)	
$flow\_related_{t-1}$				-1.140***
				(-11.78)
$flow\_residual_{t-1}$				0.0594***
				(13.02)
$\Delta inst\%_{t-1}$		0.0370***	0.0379***	0.0372***
		(23.72)	(24.06)	(23.72)
Controls	Yes	Yes	Yes	Yes
Ind FE	Yes	Yes	Yes	Yes
Qtr FE	Yes	Yes	Yes	Yes
N	175861	168746	168431	168431
R^2	0.409	0.073	0.073	0.075

Panel B: Passive investing and Price Efficiency facilitated by Short Selling demand

	Full-Sample		Down market returns		Up market returns	
	$delay\_r_t$	$delay\_coef_t$	$delay\_r\_down_t$	$delay\_coef\_down_t$	$delay\_r\_up_t$	$delay\_coef\_up_t$
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta index\%_{t-1}$	-0.538***	-0.861***	-0.131	-0.227	-0.180***	-0.309**
	(-10.03)	(-6.75)	(-1.19)	(-1.21)	(-2.69)	(-2.05)
$\Delta index\%_{t-1}$	1.522**	2.639*	-4.623***	-6.665***	0.340	-0.198
$* \, short\ interest_{t-1}$	(2.34)	(1.78)	(-3.56)	(-3.13)	(0.42)	(-0.11)
$short\ interest_{t-1}$	-0.455***	-0.591***	-0.182***	-0.181***	-0.150***	-0.232***
	(-16.21)	(-10.79)	(-5.89)	(-3.84)	(-5.70)	(-4.81)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Ind FE	YES	YES	YES	YES	YES	YES
Qtr FE	YES	YES	YES	YES	YES	YES
N	168746	168746	56308	56308	155385	155385
R^2	0.450	0.368	0.350	0.232	0.273	0.173

Panel C: Passive investing and price efficiency facilitated by Short Selling demand for Hard to Borrow stocks

	$Y = Delay\_r\_down_t$			$Y = Delay\_coef\_down_t$		
	(1)	(2)	(3)	(4)	(5)	(6)
Hard to borrow proxy	Low-arbitrage risk	Mid-arbitrage risk	High-arbitrage risk	Low-arbitrage risk	Mid-arbitrage risk	High-arbitrage risk
$\Delta index\%_{t-1}$	-0.640***	-0.404**	-0.146	-0.654*	-0.688**	-0.238
	(-3.34)	(-2.13)	(-0.73)	(-1.91)	(-2.16)	(-0.69)
$\Delta index\%_{i,t-1}$	-0.904	0.942	-4.402**	-4.067	2.071	-7.162**
$* \, short \ interest_{t-1}$	(-0.35)	(0.45)	(-2.14)	(-0.97)	(0.61)	(-2.06)
$short \ interest_{t-1}$	-0.291***	-0.156***	-0.111**	-0.260***	-0.112	-0.191***
	(-4.96)	(-3.34)	(-2.52)	(-2.84)	(-1.49)	(-2.70)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Ind FE	YES	YES	YES	YES	YES	YES
Qtr FE	YES	YES	YES	YES	YES	YES
N	18860	18866	18530	18860	18866	18530
R^2	0.378	0.324	0.265	0.272	0.207	0.155

Table 9: Subsample tests by Size and Liquidity

Table9 reports stock-level panel regression of *stock liquidity and price efficiency* on *Passive investing*, including *change of passive ownership* and *ownership weighted fund flows*, in *subsample* groups for U.S. common stocks. At beginning of each quarter, we divide sample equally into triples by *stock size* in Panel A and *stock turnover* in Panel B. We run panel regression for each sample groups, and report (x^2) and [P value] in the last column to test coefficients' differences between the smallest group and the largest group. All the controls are the same as before and not tabulated. All the independent variables are lagged by one period. Variable definitions are in Appendix A. Panel regression controls time fixed effects and industry fixed effects, and standard errors are clustered at stock level. T statistics are provided in parentheses, and *, **, *** represent significance at the 10%, 5%, and 1% confidence interval.

Panel A: Stock size

	(1)	(2)	(3)	(3) - (1)
	Small size	Mid size	Large size	Large - Small
		$\boldsymbol{Y = Spread_t}$		(x^2) [P value]
$\Delta index\%_{t-1}$	-0.0220***	-0.00311***	-0.00154***	(72.51)
	(-9.23)	(-5.84)	(-4.85)	[0.0000]
$stock\_flow_{t-1}$	-0.00282***	0.000122	0.000111*	(25.28)
	(-4.86)	(0.63)	(1.79)	[0.0000]
		$\boldsymbol{Y = Amihud_t}$		
$\Delta index\%_{t-1}$	-0.429	-0.160	-0.00721**	(0.08)
	(-0.28)	(-1.12)	(-2.21)	[0.7793]
$stock\_flow_{t-1}$	-1.070**	0.133	-0.000618	(6.60)
	(-2.57)	(1.41)	(-0.41)	[0.0102]
		$\boldsymbol{Y = Delay\_r_t}$		
$\Delta index\%_{t-1}$	-0.787***	-0.274***	-0.225***	(19.17)
	(-7.74)	(-3.67)	(-2.86)	[0.0000]
$stock\_flow_{t-1}$	-0.113***	0.0256	0.0406**	(26.10)
	(-4.65)	(1.11)	(2.30)	[0.0000]
		$\boldsymbol{Y = Delay\_coef_t}$		
$\Delta index\%_{t-1}$	-1.229***	-0.373**	-0.280	(9.80)
	(-5.42)	(-2.20)	(-1.39)	[0.0017]
$stock\_flow_{t-1}$	-0.149***	0.0504	0.126***	(17.11)
	(-2.83)	(1.02)	(3.08)	[0.0000]

Panel B: Stock liquidity

	(1)	(2)	(3)	(3) - (1)
	Illiquidity	Mid liquidity	Liquidity	Liquid-Illiquid
		$\boldsymbol{Y = Spread_t}$		(x^2) [P value]
$\Delta index\%_{t-1}$	-0.0198***	-0.00503***	-0.00273***	(39.46)
	(-7.45)	(-5.74)	(-4.81)	[0.0000]
$stock\_flow_{t-1}$	-0.00202***	-0.000977***	-0.000163	(12.05)
	(-3.93)	(-4.51)	(-1.08)	[0.0005]
		$\boldsymbol{Y = Amihud_t}$		
$\Delta index\%_{t-1}$	-3.443**	-0.0542	0.0217	(4.30)
	(-2.08)	(-0.15)	(0.20)	[0.0381]
$stock\_flow_{t-1}$	-1.005***	0.229*	-0.00404	(11.23)
	(-2.72)	(1.70)	(-0.07)	[0.0008]
		$\boldsymbol{Y = Delay\_r_t}$		
$\Delta index\%_{t-1}$	-0.560***	-0.335***	-0.421***	(1.19)
	(-5.32)	(-4.17)	(-5.94)	[0.2752]
$stock\_flow_{t-1}$	-0.0610***	-0.0398**	-0.0327*	(0.97)
	(-2.81)	(-2.01)	(-1.72)	[0.3244]
		$\boldsymbol{Y = Delay\_coef_t}$		
$\Delta index\%_{t-1}$	-1.095***	-0.384**	-0.656***	(2.27)
	(-4.52)	(-2.06)	(-4.01)	[0.1323]
$stock\_flow_{t-1}$	-0.0959**	-0.0260	-0.0414	(0.14)
	(-2.03)	(-0.56)	(-0.95)	[0.7057]

Table 10: Robustness tests

Panel A reports the impact of *passive investing* on *stock liquidity* and *price efficiency* on a *propensity score matched* sample for U.S. common stocks. The propensity scores are computed from 1:1 matching without replacement based on the covariates and interactions, with a 0.01 caliper. The treatment variable is a dummy variable equal to one if firms are in the top quartile sorted by *change of passive ownership* and *ownership weighted fund flows* in a given quarter in all panels. All regressions follow the same specifications, include stock controls, quarter fixed effects and industry fixed effects, and cluster standard errors at stock level. Variable definitions are in Appendix A.T statistics are provided in parentheses for the differences between treated group and control group, and *, **, *** represent significance at the 10%, 5%, and 1% confidence interval.

Panel A: Propensity score matching

	(1)	(2)	(3)	(4)
	$Spread_t$	$Amihud_t$	$Delay\_r_t$	$Delay\_coef_t$
$\Delta \boldsymbol{Index}\%_{\boldsymbol{t-1}}$				
$Treated$	0.0048	0.5892	0.3318	1.5764
$Control$	0.0049	0.5553	0.3449	1.5955
$Difference$	-0.0001	0.0340	-0.0131***	-0.0191***
	(-0.83)	(1.46)	(-6.45)	(-4.40)
obs	83696	83696	83696	83696
$\boldsymbol{Stock\_flow_{t-1}}$				
$Treated$	0.0045	0.4949	0.3284	1.5698
$Control$	0.0047	0.4689	0.3460	1.5980
$Difference$	-0.0002**	0.0260	-0.0176***	-0.0282***
	(-2.33)	(1.24)	(-8.63)	(-6.44)
obs	81980	81980	81980	81980

Panel B

Panel B regresses price delay on index ownership level and stock controls levels in column (1), and regresses change of price delay on change of index ownership and changes of stock controls in column (2) and (3). Regression settings are the same as before.

	$Y = Delay\_r_t$		$Y = \Delta Delay\_r_t$	
	(1)		(2)	(3)
		$stock\_flow_{t-1}$		-0.0705***
				(-6.20)
$index\%_{t-1}$	-1.009***	$\Delta index\%_{t-1}$	-0.434***	
	(-34.25)		(-6.64)	
$inst\%_{t-1}$	-0.0543***	$\Delta inst\%_{t-1}$	-0.0277**	-0.0349***
	(-10.33)		(-2.24)	(-2.82)
$size_{t-1}$(log)	-0.0487***	$\Delta size_{t-1}$(log)	-0.0683***	-0.0656***
	(-44.44)		(-10.05)	(-9.59)
btm_{t-1}	-0.00742***	Δbtm_{t-1}	0.00125	0.00103
	(-21.14)		(1.01)	(0.83)
$std(ret)_{t-1}$	-1.091***	$\Delta std(ret)_{t-1}$	-0.530***	-0.518***
	(-15.35)		(-6.98)	(-6.82)
$turnover_{t-1}$	-0.419***	$\Delta turnover_{t-1}$	-1.323***	-1.337***
	(-3.19)		(-6.90)	(-6.95)
$price_{t-1}$(log)	-0.0155***	$\Delta price_{t-1}$(log)	0.00709	0.00508
	(-10.27)		(1.02)	(0.73)
$analyst_{t-1}$(log)	-0.00487***	$\Delta analyst_{t-1}$(log	-0.00321	-0.00340
	(-3.10)		(-1.24)	(-1.31)
dv_{t-1}	0.333***	Δdv_{t-1}	-0.462***	-0.463***
	(88.99)		(-29.44)	(-29.75)
Intercept	0.605***	*Intercept*	-0.166***	-0.0256*
	(21.09)		(-11.38)	(-1.92)
Ind FE	YES	*Ind FE*	YES	YES
Time FE	YES	*Time FE*	YES	YES
N	175749	*N*	168638	168325
R^2	0.495	R^2	0.267	0.267

Panel C

Panel C uses panel regression in column (1), controlling firm fixed effects and time fixed effects and clustering standard errors at stock level, and uses quarterly Fama-Macbeth regression in column (2).

	$\boldsymbol{Y = Delay\_r_t}$	
	Firm FE &Qtr FE	Fama-Macbeth
	(1)	(2)
$\Delta index\%_{t-1}$	-0.309***	-0.963**
	(-7.01)	(-2.06)
$\Delta inst\%_{t-1}$	-0.0160	-0.0346
	(-1.50)	(-1.55)
$size_{t-1}$(log)	-0.0671***	-0.0465***
	(-17.94)	(-19.90)
btm_{t-1}	0.000704	-0.00319***
	(0.84)	(-4.03)
$std(ret)_{t-1}$	-1.072***	-0.602***
	(-13.73)	(-3.37)
$turnover_{t-1}$	-0.606***	-0.976***
	(-4.03)	(-4.27)
$price_{t-1}$(log)	-0.0188***	-0.0204***
	(-4.89)	(-10.58)
$analyst_{t-1}$(log)	-0.0131***	-0.0105***
	(-6.16)	(-4.67)
dv_{t-1}	0.140***	0.364***
	(36.09)	(28.52)
Intercept	0.799***	0.657***
	(40.81)	(34.76)
N	168657	168657
R^2	0.572	0.437

Panel D

Panel D reports cross-sectional averages of time-series Granger causality regression. We construct first-differences in all variables, use change of price efficiency as dependent variable in column (1) and change of index ownership as dependent variable in column (2).

	$\boldsymbol{Y = \Delta Delay\_r_t}$	$\boldsymbol{Y = \Delta Index\%_t}$
	(1)	(2)
$\Delta index\%_{t-1}$	-0.612***	-0.373***
	(-2.88)	(-7.96)
$\Delta delay\_r_{t-1}$	-0.460***	0.0000590
	(-47.83)	(0.17)
$\Delta inst\%_{,t-1}$	-0.0219	0.00851***
	(-1.19)	(3.06)
$\Delta size_{t-1}$(log)	-0.0706***	-0.00102
	(-5.41)	(-1.34)
Δbtm_{t-1}	0.0513	0.000663
	(1.28)	(0.40)
$\Delta std(ret)_{t-1}$	-0.394***	-0.0200***
	(-3.45)	(-3.63)
$\Delta turnover_{t-1}$	-1.670***	0.000152
	(-4.71)	(0.01)
$\Delta price_{t-1}$(log)	-0.00433	0.00207***
	(-0.39)	(2.80)
$\Delta analyst_{,t-1}$(log)	-0.00346	0.000592***
	(-1.18)	(3.73)
Intercept	0.00343	0.0000116
	(0.41)	(0.01)
N	168657	168657
R^2	0.208	0.207

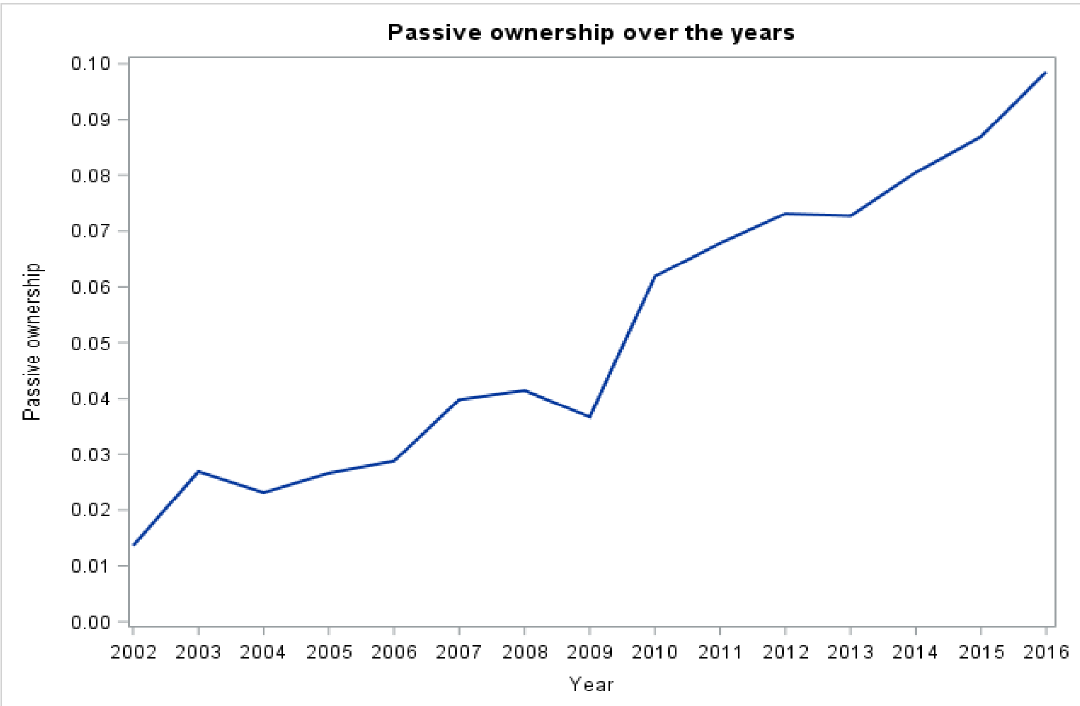
Panel E

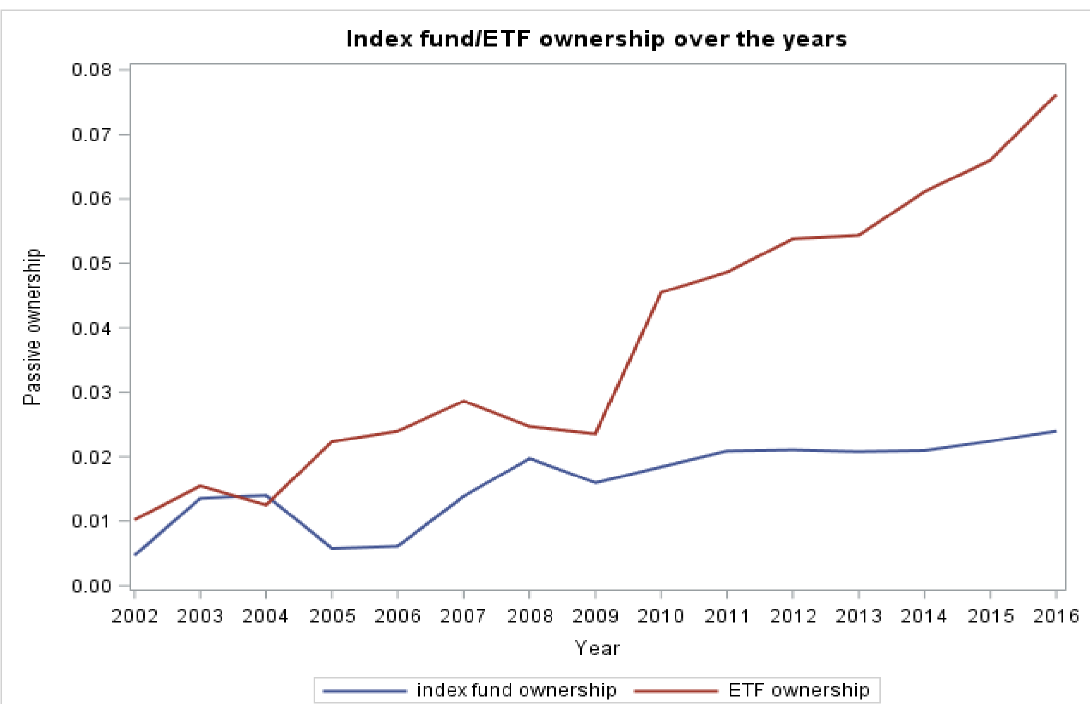
Panel E divides sample periods by year before(after) 2008 and CBOE's volatility index (VIX).

	$\boldsymbol{Y = Delay\_r_t}$			
	Year>=2008	Year<2008	VIX>=18	VIX<18
	(1)	(2)	(3)	(4)
$\Delta index\%_{t-1}$	-0.388***	-0.571***	-0.440***	-0.457***
	(-6.92)	(-6.46)	(-7.12)	(-5.98)
$\Delta inst\%_{,t-1}$	-0.00711	-0.0653***	-0.0208	-0.0485***
	(-0.51)	(-3.82)	(-1.36)	(-3.20)
$size_{,t-1}$(log)	-0.0405***	-0.0505***	-0.0437***	-0.0435***
	(-32.42)	(-29.35)	(-33.54)	(-31.53)
btm_{t-1}	-0.00335***	-0.00549***	-0.00403***	-0.00397***
	(-8.68)	(-13.74)	(-9.54)	(-11.46)
$std(ret)_{t-1}$	-0.645***	-1.334***	-1.158***	-0.510***
	(-7.16)	(-10.95)	(-13.37)	(-4.58)
$turnover_{t-1}$	-1.120***	-1.560***	-1.468***	-0.907***
	(-7.34)	(-7.42)	(-9.57)	(-5.28)
$price_{t-1}$(log)	-0.0210***	-0.0248***	-0.0182***	-0.0255***
	(-11.85)	(-9.92)	(-9.78)	(-13.81)
$analyst_{t-1}$(log)	-0.0155***	-0.0188***	-0.0168***	-0.0178***
	(-8.79)	(-7.36)	(-9.31)	(-8.70)
dv_{t-1}	0.379***	0.355***	0.397***	0.353***
	(79.81)	(64.63)	(78.26)	(75.13)
Intercept	0.569***	0.747***	0.513***	0.673***
	(54.09)	(20.07)	(18.30)	(32.46)
N	101938	66700	75994	92644
R^2	0.480	0.455	0.528	0.432

Figure 1: Index fund ownership by year

Figure1 plots averaged index fund ownership from 2002 to 2016 for U.S. common stocks. Ownership is calculated as percentage of stock outstanding shares held by index funds at the end of each quarter.





Appendix A: Variable Definition

Variable	Definition
Index%	Percentage of stock shares outstanding held by passive funds at the end of quarter. Data is from Thomson Reuters Mutual Fund holdings (S12)
Num_index	Number of passive funds holding that stock
Inst%	Percentage of stock shares outstanding held by institutions at the end of quarter. Limit the percentage between 0 and 100%. Data is from Thomson Reuters Institutional holdings (S34)
Num_inst	Number of institutions holding that stock
Stock_flow	Ownership weighted average flows of passive funds holding that stock. Fund flow is calculated as $(\frac{TNA_{j,t}-TNA_{j,t-1}*(1+r_q)}{TNA_{j,t-1}})$ each quarter, and data is from CRSP Mutual Funds. Weight is ownership of passive funds holding that stock in last period.
Flow_related	Fitted values from quarterly regressions of changes of index ownership: ΔIndex% on Stock_flow, controlling quarter fixed effects and clustering standard errors at individual firm level
Flow_residual	Residual values from quarterly regressions of changes of index ownership: ΔIndex% on Stock_flow, controlling quarter fixed effects and clustering standard errors at individual firm level
Stock_ETF flow	Ownership weighted average flows of ETFs holding that stock. We filter ETF sample which have a share code of '73' in CRSP or CRSP ETF flag "F". Weight is ownership of ETFs holding that stock in last period.
Stock_index fund flow	Ownership weighted average flows of index mutual funds holding that stock. We filter index mutual funds sample which are not included in the ETF sample but in total index sample. Weight is ownership of index mutual funds holding that stock in last period.
Amihud	Quarterly average of daily Amihud illiquidity ratio, where daily Amihud is daily absolute return over dollar volume
Analyst	Number of analysts in that quarter. Data is from IBES. Missing values are set to zero
Arbi_risk	Mean squared error of residuals (RMSEs) from Carhart's four-factor model using daily stock returns within a quarter with a minimum of 50 daily observations
Beta	Coefficient of stock's daily excess returns on daily market excess returns within a quarter with a minimum of 50 daily observations
Book to market	Book value for the fiscal year ended before the most recent June 30, divided by market capitalization of December 31 during that fiscal year
Delay_r	Ratio of R^2 from restricted market model and unrestricted market model. Unrestricted market model is to regress stock returns on contemporaneous and five lags of lagged market returns over each quarter with a minimum of 50 daily observations; Restricted market model limits the coefficients of lagged market returns to zero.
Delay_coef	Ratio of lag-weighted sum of coefficients of lagged market returns relative to the sum of all coefficients, scaled by standard errors of the coefficients
Delay_r_down	Use only negative contemporaneous and lagged market returns in the above restricted model and unrestricted model with a minimum of 30 daily observations. Delay_coef_down is similarly defined.
Delay_r_up	Use only positive contemporaneous and lagged market returns in the above restricted model and unrestricted model with a minimum of 30 daily observations. Delay_coef_up is similarly defined

Price	Quarterly close price, where price is limited to $1 to $1000
Short interest	Average monthly short interest during quarter, where monthly short interest is shares sold short divided by stock shares outstanding. Data is from Compustat and CRSP
Size	Market capitalization as price times shares outstanding, expressed in millions
Spread	Difference between ask and bid price divided by daily closed price, averaged each quarter.
Std(ret)	Standard deviation of stock daily returns within a quarter
Sync_mkt	$\log(\frac{R^2}{1-R^2})$, where R^2 is from the regression of stocks' daily excess returns on daily market excess returns within a quarter with a minimum of 50 daily observations
Sync_ind	$\log(\frac{R^2}{1-R^2})$, where R^2 is from the regression of stocks' daily returns on daily market returns and industry returns within a quarter with a minimum of 50 daily observations. Industry returns are defined as two-digit SIC industry returns
Turnover	Quarterly average of daily turnover, where turnover is stock trading volume over stock shares outstanding
|1-vr(1,5)|	Absolute value of difference between one and the ratio of weekly stock return variance to five times daily stock return variance.

Chapter 4

Long-term Index Fund Ownership and Stock Returns

Abstract

We examine the implications of stock ownership by index funds for shareholder value. Consistent with recent findings that stock ownership by passive funds contributes to improved governance, we document a strong positive relation between the duration of passive fund holdings and subsequent stock performance. This positive relation is more pronounced for firms with recent poor performance, and for smaller firms and firms with higher allocation weights in passive funds' portfolios. Our results support the view that index funds, although passive in their investment decisions, successfully contribute to long-term value creation by actively engaging with firms on matters of governance.

4.1. Introduction

There has been a dramatic growth in the assets of passively managed index mutual funds in recent years. For example, according to the Investment Companies Institute (ICI), domestic equity index mutual funds and exchange traded funds (ETFs) received $1.2 trillion in net new cash, including reinvested dividends, between 2007 and 2015. In stark contrast, actively managed domestic equity mutual funds experienced net outflows of $835 billion (even after accounting for reinvested dividends) over the same period. As of the end of 2015, domestic index fund assets accounted for about 35% of total assets held by equity mutual funds.

Not surprisingly, the growing importance of passive institutional investors such as index mutual funds has been the focus of much interest and has sparked to considerable debate regarding their impact on firm-level governance. In a recent paper, Appel, Gormley, and Keim (2016) examine the role of passive mutual fund companies in corporate governance and find that such investors are not merely passive owners. In particular, they find that passive investors appear to play an important role in pushing their portfolio companies to adopt shareholder-friendly policies, including an increase in the number of independent directors and the elimination of poison pills and dual-class share structures. More generally, the authors document that passive ownership is associated with a decline in shareholder support for management proposals and an increase in support for governance-related shareholder proposals. Furthermore, longer-term passive stock ownership is associated with significant improvements in the firm's return on assets and Tobin's Q. These results are broadly consistent with the conclusions of earlier studies that found that institutional investors, including those that index a large portion of their portfolios, can affect corporate behavior (e.g., Carleton, Nelson, and Weisbach, 1998; Del Guercio and Hawkins, 1999; Gillan and Starks, 2000; Harford, Kecskés, and Mansi, 2018).

Motivated by recent results in the literature, in this paper we examine whether improvements in firm-level governance due to long-term passive ownership by index funds lead to improved returns to investors in the affected firms. Unlike actively managed funds, index funds do not have discretion over which stocks to hold, and in particular, they do not have the option of selling stocks that underperform. Hence, it could be argued that index funds have a stronger incentive to undertake improvements in the governance of their portfolio firms.[14] As F. William McNabb III, Vanguard's Chairman and CEO, wrote in a recent letter[15] to the boards of directors of Vanguard funds' largest portfolio holdings,

We are large, we don't make a lot of noise, we are focused on the long term, and we don't tend to rush into and out of investments. In the past, some have mistakenly assumed that our predominantly passive management style suggests a passive attitude with respect to corporate governance. Nothing could be further from the truth. We want to see our clients' investments grow over the long term, and good governance is a key to helping companies maximize their returns to shareholders.

If index funds' efforts in improving firm governance and long-term value are effective, it is reasonable to expect that their substantial holdings will have an impact on stock performance as well. We provide direct evidence on this important issue in this paper.

We identify a sample of U.S. passive and active equity funds during the period 2003:Q1 to 2015:Q3. The sample includes 608 funds classified as passive equity funds, including index

---

[14] For example, according to the Global Governance Principles adopted by the largest U.S. public pension fund, CalPERS, which has a substantial allocation to indexed portfolio investments, "CalPERS prefers constructive engagement to divesting as a means of affecting the conduct of entities in which it invests. Investors that divest lose their ability as shareowners to influence the company to act responsibly." (Source: CalPERS Global Governance Principles, Updated: March 16, 2015, p. 9)

[15] https://about.vanguard.com/vanguard-proxy-voting/CEO_Letter_03_02_ext.pdf

mutual funds and ETFs. We obtain data on stocks held by funds using the Thomson-Reuters mutual fund holdings (S12) database. It is likely that passive funds' impact on governance would be stronger in the case of stocks they hold for a long time. Accordingly, at the end of every quarter we construct a measure of the duration of ownership of each stock by every fund during the previous 20 quarters, following Cremers and Pareek (2016). We then average this measure across all passive funds to construct an overall duration measure for each stock. For each stock, this measure reflects the weighted duration of the investment in the stock by all passive funds.

Our key hypothesis links the strength of monitoring by passive fund investors, as reflected in the duration of holdings measure, to future stock returns. In tests based on cross-sectional regressions, we find that our passive fund stock holding duration measure is significantly and positively related to future raw and excess returns at horizons up to 24 months. For example, the results imply that a one-standard-deviation increase in the (log of the) passive funds' stock holding duration measure for a particular stock is associated with an increase in the stock's quarterly return by 48 basis points over the next 3 months. The corresponding increase in the stock's annual return is 189 basis points over the next 12 months, and 161 basis points during the second year. Our results are qualitatively similar when using an alternative measure for duration of passive funds' stock holdings based on the funds' portfolio turnover (see, for example, Gaspar, Massa, and Matos, 2005). Interestingly, we find that a similar stock holding duration measure based on the portfolio holdings of actively managed funds has much weaker predictive ability for future stock returns. Specifically, in predictive return regressions, the average coefficients on the active funds' (excluding closet indexers) holdings duration measures are 0.308 and 0.759 for next quarter returns and next year returns, and are only marginally significant at the 10% level for next quarter returns. The coefficients decline in magnitude to 0.247 and 0.492 and become statistically insignificant when the passive funds' holdings duration measure is included as a control. We also show that our results are not driven by closet indexers: After controlling for passive funds' holding duration, closet indexers have a limited role in predicting returns.

Next, we adopt a portfolio approach and sort funds into quintile portfolios according to the duration measure based on passive funds' stock holdings. A spread portfolio that is long the longest duration fund portfolio and short the shortest duration fund portfolio earns a monthly 4-factor (three Fama-French factors and the Carhart momentum factor) alpha of 70.9 basis points (or 8.5% annually) and a 5-factor (four factors plus the Pastor and Stambaugh liquidity factor) alpha that equals 70.5 basis points (or 8.46% annually) during the period 2003:Q1 to 2015:Q3.

What explains the predictive ability of our measure of the duration of passive funds' stock holdings? To explore this issue further, we split the sample of stocks based on their performance during the previous 12 months and 36 months. If the predictive ability is indeed driven by the improvements in firm-level governance brought about by long-term ownership by passive funds, we would expect a stronger positive relation between the duration of holdings measure and future stock returns for the worst performing stocks. Similarly, we would expect this relation to be stronger for stocks with smaller market capitalization, which may be more susceptible to the influence of passive fund owners, especially when they are large. We also expect a stronger relation between the duration of holdings measure and future stock returns during periods in which the market is more volatile, when passive funds are likely able to exert a stronger influence on management.

Our results based on cross-sectional tests provide support for each of these three predictions. The predictive ability of the duration of holdings measure for future returns at the 3-month, 12-month, and 24-month horizon is stronger for the worst performing stocks (i.e., stocks with below-median performance during the past 12 months or past 36 months). In addition, the predictive ability of the passive fund stock holdings' duration measure is stronger for smaller firms, i.e., for firms with below-median market capitalization. The measure's predictive ability is also more pronounced during more volatile market periods.

As a further test of the importance of the monitoring role played by passive funds and its impact on future stock returns, we include in our test design a control variable that is a measure of the passive funds' aggregate allocation weight to a particular stock. In cross-sectional tests, the interaction term involving the allocation weight-based variable and the duration of holdings measure is significantly positively related to future stock returns. The relationship is positive and significant at multiple horizons up to 1 year ahead. These results suggest that the passive funds' allocation weight has significant, marginal predictive power for stock returns at both short and long horizons.

Finally, we compare a subsample of stocks that rank at the bottom among stocks in the Russell 1000 index, based on market capitalization, to those that rank near the top among stocks in the Russell 2000 index. Stocks near the boundary of the index membership cutoff are likely to be quite similar in their characteristics, with one important exception. Since index funds' stock allocations are based on market valuations, stocks at the top of the Russell 2000 index would be weighted more heavily in portfolios of index funds (targeting the Russell 2000 index). On the other hand, stocks near the bottom of the Russell 1000 index will be featured less prominently in portfolios of index funds (targeting the Russell 1000 index). This distinction allows us to perform a relatively clean test of the impact of passive fund ownership on stock returns. We find that the predictive ability of passive funds' holdings duration measure for future stock returns is much stronger for stocks at the top of the Russell 2000 index compared to those at the bottom of the

Russell 1000 index. This finding is consistent with the idea that significant holdings of passive funds are associated with more effective monitoring by the funds.

We rule out the possibility that our results are driven solely by the potential persistent buying-related price pressure experienced by stocks that are constituents of various market indexes. In particular, our results are robust to controls for lagged stocks returns that proxy for past asset flows. Furthermore, we confirm the predictive ability of the holdings’ duration measure at longer horizons up to 2 years.

We also address concerns that reverse causality can explain our results. Under this explanation, the better performing stocks would mechanically enjoy a longer duration of holdings. To explore this possibility, we examine the correlation between the duration of holdings measure and past stock returns. We find that the correlation is in fact quite weak.

Our paper also contributes to the literature on the duration of fund holdings or trade frequency, and fund performance. In this context, Cremers and Pareek (2016) document that among active funds with high active share, the funds that trade infrequently tend to outperform on average by about 2%per year. Furthermore, among funds with long holding durations, the high active share funds outperform the low active share funds. They attribute their results to the ability of a subset of skilled active fund managers who are better at identifying instances of security mispricing that are eventually corrected over the long term.

Our results are also consistent with the findings of Harford, et al., (2018), who document the favorable impact of long-term investors on shareholder returns. In contrast to Harford et al. (2018), whose primary focus is on the impact of investor horizons on corporate decisions, our analysis specifically focuses on the implications of stock ownership by index funds.[16] Hence, in our analysis we directly identify passive fund investors, namely, index funds and ETFs, rather than relying on a noisy activeness measure (e.g., the active share) in order to classify investors as being active or passive. It should also be noted that index funds are long-term investors by design, and their investors are more likely to be patient compared to investors in active funds. Since index funds do not engage in active security selection, our results suggest that the monitoring role of passive investors is the most likely explanation for our findings. On the other hand, closet indexers (identified using the active share measure) have considerable flexibility and discretion in their investment choices. Their motivations and investment constraints can be quite different from that of genuine index funds. Indeed, our results show that after controlling for the effect of index funds and ETFs, duration measures related to stock ownership by either closet indexers or active funds with long-term horizons are unrelated to future stock returns.

The rest of the paper is organized as follows. Section 2 develops our main testable hypotheses. Section 3 discusses the data, sample, and variable construction. Section 4 presents the main findings on the effect of passive funds' long-term investments on stock performance and discusses tests of various hypotheses. Section 5 compares passive funds and active funds, and Section 6 concludes.

4.2. Testable hypotheses

Appel, Gormley, and Keim (2016) show that passive funds are not merely passive owners; instead, they play an important role in firm governance. If passive funds have the incentive to

[16] As Harford et al. (2018) acknowledge, "Itis not our objective to study the consequences of indexing as such." (p. 429). In order to establish a causal link between investor horizons and corporate outcomes, for part of the analysis they classify investors into "indexers" and "non-indexers." This classification is based on an active share measure calculated for the institutional investors in their sample.

monitor firms in their portfolios and are effective in improving the firms' governance, their long-term holdings should favorably impact their stock performance relative to firms that are not in their portfolio. As we discuss in the introduction (and subsequently in more detail in the data section), we use a stock's passive holdings duration (churn ratio) to measure passive funds' long-term commitment to the stock. We therefore propose the following hypothesis.

H1: A stock's passive holdings duration (churn ratio) positively (negatively) predicts future returns.

As "permanent" shareholders, passive funds do not have the option of selling their positions in underperforming stocks. Hence, we expect that they have a stronger incentive to monitor and influence firms that have been doing poorly in the past. This suggests the following hypothesis.

H2: The return predictability of the passive holdings' duration is stronger for underperforming stocks.

Given their monitoring incentives, passive funds' ability to influence the firm depends on the size of the firm. Everything else equal, we expect that passive funds have a stronger impact on small firm performance. Further, we expect that the passive funds' monitoring incentive would be stronger during more volatile periods, when there is greater uncertainty about the performance of the stocks they invest in. Hence, we have the following hypothesis.

H3: The return predictability of the passive holdings' duration is stronger for smaller firms, and during more volatile market conditions.

Many passive funds invest in hundreds of stocks. Despite the resources available to large funds, it would be difficult for them to pay equal attention to all stocks. Hence, we expect that passive funds would be more effective in monitoring stocks that have greater weights in their holdings (adjusting for the market weight of the stocks). Accordingly, we have the following hypothesis.

H4: The return predictability of passive funds holdings duration is stronger for stocks with greater excess weights (relative to market value weights) in passive funds' portfolios.

Recent literature has argued that active funds' long-term holdings outperform passive funds' holdings. We note that by their design, passive funds are long-term investors and have incentives to monitor and influence firm governance and performance. If their monitoring is effective and favorably impacts stock returns, it is possible that some "long-term" active funds mimic passive funds' long-term investment holdings. This suggests the following hypothesis.

H5: Controlling for the passive funds' holdings duration effect, the return predictability of the active funds' holdings duration is diminished.

4.3. Data and sample construction

4.3.1. Passive and active funds sample construction

Our data for U.S. mutual funds comes from the CRSP Survivor Bias Free U.S. mutual fund database and Thomson Reuters mutual fund holdings (S12) database linked by MFLINKS. We exclude bond funds and international funds to ensure that only domestic equity mutual funds are left in the sample. Additionally, we require that equity funds in our sample have allocations to common stocks between 80% and 105%, hold at least 10 stocks, and manage assets in excess of $5 million. Since fund characteristics provided by CRSP are at the share class level, we calculate value-weighted fund characteristics, such as turnover ratio, across multiple share classes within a fund using total net assets (TNA) as weights. Finally, we require that funds in our sample have available shareholding information and have at least 1year's worth of holdings history.

To classify funds as either passively managed funds or actively managed funds, we examine the CRSP index fund/ETF indicators. To identify passive funds that are not explicitly identified by these indicators, we follow Appel, et al.(2016) and screen the remaining sample using keywords in their names.[17] The remaining funds in our sample are classified as active funds. This procedure yields 608 passive funds and 2,732 active funds over the period from first quarter of 2003 to third quarter of 2015. We use 2003 as the starting year for the sample since there are substantially fewer passive funds prior to this year. We compute the percentage of stocks' shares outstanding owned by passive funds (*index*%) and by active funds (*active*%) at the end of each quarter.

4.3.2. Stock long-term ownership by passive (active) funds

We focus on U.S. common stocks (share code 10 or 11) that are listed on NYSE, AMEX, or NASDAQ from 2003.Q1 to 2015.Q3. We eliminate stocks with prices below $1 or above $1,000. Further, we require that a stock be held by a fund for at least 2 consequent quarters to exclude the occasional addition (removal) of stocks into (out of) funds.

We construct two measures of funds' (long-term) investment on a stock. The first measure is the stock-level "duration" measure, as motivated by Cremers and Pareek (2016). By tracing back the holding periods and weighting the buys and sells in each period, this measure captures how

[17] The Strings we use to identify index funds are: Index, Idx, Indx, Ind_, Russell, S&P, S & P, SandP, SP, DOW, Dow, DJ, MSCI, Bloomberg, KBW, NASDAQ, NYSE, STOXX, FTSE, Wilshire, Morningstar, 100, 400, 500, 600, 900, 1000, 1500, 2000, 5000, ishares, powershares, SPDR, QQQ, ETF, EXCHANGE TRADED, EXCHANGE-TRADED, PROFUNDS, SPA MG, MARKET GRADER.

long a stock has continuously been held by one fund at a particular time. Specifically, at the end of quarter *t*, the holding duration of stock *i* in passive fund *j* is given by:

$$Duration_{i,j,T} = \sum_{t=T-W+1}^{T} \left(\frac{(T-t)\alpha_{i,j,t}}{H_{i,j,T-w} + B_{i,j}} \right) + \frac{W * H_{i,j,T-W}}{H_{i,j,T-w} + B_{i,j}}, \quad (1)$$

where $\alpha_{i,j,t}$ is change in percentage of shares outstanding of stock *i* held by index *j* between quarter *t*-1 and quarter *t*, and $\alpha_{i,j,t} > 0$ for buys and $\alpha_{i,j,t} < 0$ for sells. The term $H_{ij,T-W}$ is the percentage of total shares outstanding of stock *i* held by fund *j* at the end of quarter T-W; and $B_{i,j}$ is percentage of total shares outstanding of stock *i* bought by fund *j* during time period between quarters T-W and T. Consistent with the literature, we choose w=20, since any trading prior to 5 years ago would not be as relevant when assessing holding decisions in year 0. Next, we compute stock duration across all passive funds by either equally weighting $Duration_{i,j,T}$ or averaging $Duration_{i,j,T}$ using passive fund ownership of the stock *index* % as the weight across all passive funds that hold the stock:

$$Dur - equal_{i,T} = \frac{\sum_j Duration_{i,j,T}}{N_j} \quad (2)$$

$$Dur - weighted_{i,T} = \frac{\sum_j Duration_{i,j,T} * Index\%_{i,j,T}}{\sum_j Index\%_{i,j,T}} \quad (3)$$

Similarly, we construct the duration measures based on active funds $Duration - ac_{i,j,T}$. We next compute stock duration in active funds by equally weighting $Duration - ac_{i,j,T}$ or averaging $\text{Duration-ac}_{i,j,T}$ using active fund ownership of the stock *active*% as the weight across all active funds that hold the stock:

$$Dur - equal - ac_{i,T} = \frac{\sum_j Duration - ac_{i,j,T}}{N_j} \tag{4}$$

$$Dur - weighted - ac_{i,T} = \frac{\sum_j Duration - ac_{i,j,T} * Active\%_{i,j,T}}{\sum_j Active\%_{i,j,T}} \tag{5}$$

The second long-term fund investment measure we consider is the churn ratio. The churn ratio has been widely used to proxy for fund investment horizon (see, for example, Gaspar, et al., 2005; Yan and Zhang, 2009; Cella, Ellul, and Giannetti, 2013). Instead of focusing on the fund level churn ratio, we measure the average churn ratio across passive funds for a stock, as follows:

First, the turnover of stock *i* by passive fund *j* in quarter *t* is given by:

$$CR_{i,j,t} = \frac{|N_{i,j,t}P_{i,t} - N_{i,j,t-1}P_{i,t-1} - N_{i,j,t-1}\Delta P_{i,t}|}{\frac{N_{i,j,t}P_{i,t} + N_{i,j,t-1}P_{i,t-1}}{2}}, \tag{6}$$

where $P_{i,t}$ and $N_{i,j,t}$ are price and number of shares of stock *i* held by passive fund *j* at the end of quarter *t*. We then calculate the churn ratio of passive fund *j* for stock *i* by averaging across the prior 4 quarters:

$$CR_{i,j,t(r)} = \frac{1}{4} \sum_{r=1}^{4} CR_{i,j,t-r+1} \tag{7}$$

Similarly, we calculate stock-level churn ratio by equally averaging $CR_{i,j,t(r)}$ or averaging $CR_{i,j,t(r)}$ using passive fund ownership of stock *i* as the weight across all passive funds holding that stock:

$$CR - equal_{i,t} = \frac{\sum_j CR_{i,j,t(r)}}{N_j} \tag{8}$$

$$CR - weighted_{i,t} = \frac{\sum_j CR_{i,j,t(r)} * Index\%_{i,j,t}}{\sum_j Index\%_{i,j,t}} \tag{9}$$

We also select active funds and calculate stock-level churn ratio in active funds, by equally averaging $CR - ac_{i,j,t(r)}$ or averaging $CR - ac_{i,j,t(r)}$ using active fund ownership of stock i as the weight across all active funds holding that stock:

$$CR - equal - ac_{i,t} = \frac{\sum_j CR - ac_{i,j,t(r)}}{N_j} \tag{10}$$

$$CR - weighted - ac_{i,t} = \frac{\sum_j CR - ac_{i,j,t(r)} * Active\%_{i,j,t}}{\sum_j Active\%_{i,j,t}} \tag{11}$$

Panel A of Table 1 reports the time-series mean, standard deviation, minimum, median, and maximum values of the cross-sectional averages of duration, churn ratio, and proportional stock ownership for passive funds and active funds across 51 quarters. The holdings duration's measure is winsorized at the 1st percentiles and expressed in number of quarters. The churn ratio and ownership measure are winsorized at the 1st and 99th percentiles.

Stocks in passive funds have an ownership- (equal-) weighted average holding duration of 7.910 quarters (7.118 quarters). In comparison, stocks have an ownership- (equal-) weighted average duration of 5.785 quarters (5.289 quarters) in active funds, which suggests that passive funds tend to hold stocks for relatively longer periods. Similarly, stocks in passive funds have a smaller churn ratio compared with those in active funds. On average, index funds own 4.7% of the outstanding shares of stocks they invest in. As expected, active funds hold less diversified portfolios, and on average, they hold a larger proportion of the shares of stocks they own with a mean of 10.3%. Panel B of Table 1 reports the time-series averages of cross-sectional correlations between duration, churn ratio, and index ownership. As expected, the correlation between duration and churn ratio is negative and equals -0.434 for the ownership-weighted measures, and -0.38 for equal-weighted measures. Moreover, ownership- (equal-) weighted duration is positively related to index ownership at 25.3% (21.9%), but as expected, the churn ratio measures display much lower correlations with index ownership. We also calculate fund level churn ratio and duration for passive funds, and the correlation is negative and equals -0.625.

Figures 1, 2, and 3 depict the time-series trends for our key variables. Figure 1 shows that passive ownership increases over the years, from around 2% in early 2003 to over 8% in late 2015. The obvious increase occurs in late 2008, which coincides with the growing importance of passive funds, especially following the global financial crisis. In contrast, active ownership is relatively stable at around 10% during the sample period, except for a decrease during late 2008. Figures 2 and 3 compare stock holding's duration and stock churn ratio for passive funds and active funds, respectively. Duration is slightly increasing and the churn ratio is more volatile but decreases over time, suggesting that in general, stocks tend to be held by funds for longer than before. Second, passive duration is always larger than active duration, and passive churn ratio is always smaller than active churn ratio. Third, during the financial crisis there is a decline in passive duration and an obvious increase in passive churn ratio.

4.3.3. Measures of relative importance of a stock in passive fund holdings

In some of our tests, we examine the relative importance of a portfolio weight of a stock held by passive funds for the funds' monitoring incentive. If passive funds overweight a particular stock relative to the stock's weight in the market portfolio, we would expect the funds to have a stronger incentive to monitor the stock.

Accordingly, we construct an excess weight measure for stock *i* at the end of quarter *t*:

$$Excess\ Weight_{i,t} = w_{it} - \overline{w_{\iota t}} \tag{12}$$

where $w_{i,t}$ is weight of stock *i* in overall passive fund holdings, and $\overline{w_{i,t}}$ is the weight of stock *i* in the market portfolios. We use the value-weighted portfolio of the U.S. domestic equity stocks in our sample as a proxy for the market portfolio. We then sort all stocks in our sample each quarter into halves based on the excess weight measure, and define a dummy variable *important*, which equals one if a stock's excess weight is above the cross-sectional median value, and 0 otherwise.

4.3.4 Additional stock characteristics as regression control variable.

In the subsequent regressions, we include the following stock characteristics as control variables:

Price: share price from CRSP. We exclude stocks priced below $1 or above $1,000.

Size: stock market capitalization in millions.

Btm: book to market, book value for the fiscal year ended before the most recent June 30, divided by market capitalization of December 31 during that fiscal year using data from Compustat and CRSP. Btm is winsorized at the 1st and 99th percentiles.

Volatility: standard deviation of monthly returns over the previous 2 years.

Turnover: average monthly traded shares divided by shares outstanding, calculated over the previous3 months.

Age (months): number of months since first returns appear in CRSP

Beta: market beta calculated each quarter by regressing a stock's daily excess return on the daily market excess return during the quarter.

SP_500 dummy: dummy variable for S&P 500 index membership.

Ret(t, t-3): cumulative gross returns over the past 3 months.

Ret(t-3, t-12): cumulative gross returns over the 9 months preceding beginning of the filing quarter.

Ret(t-12, t-36): cumulative gross returns over 2 years before the last year.

All of the variables except returns are measured quarterly. Panel A of Table 1 shows that there are on average 3,726 stocks every quarter. The average firm has a stock price of $27.157, a market capitalization of $4.44 million, a book-to-market ratio of 1.107, and a beta of 1.025. Average stock volatility and turnover are 12.1% and 17.1%, respectively. Panel B of the table shows that duration (churn ratio) has strong positive (negative) correlation with firm age and the SP500 dummy, and negative (positive) correlation with volatility and turnover, suggesting that these two measures can capture stock-level holding horizons in funds. Following Yan and Zhang (2009), we express all variables in natural logarithms with the exception of stock returns, beta, S&P500 dummy, and churn ratio.

4.4 Return predictability of the long-term investment of passive funds

4.4.1 Long-term passive fund ownership and future stock returns

Appel et al. (2016) show that passive funds ownership affects firms' governance and investment decisions. If the monitoring role of passive funds were effective, their long-term ownership could have positive impacts on stock performance, everything else equal. We formally test this hypothesis (H1 in Section 2) in this subsection. We measure passive funds' long-term ownership by duration and churn ratio, both defined in Section 3. We measure future stock returns with three holding periods: return for the next 3 months (3-month-ahead return), return for the next 12 months (1-year-ahead return), and return from the end of month 12 to month 24 (2-year-ahead return). Each quarter, for each future return measure, we conduct cross-sectional regressions of future returns on the passive duration (churn ratio) measure, controlling for other firm characteristics. Firm-level control variables include stock market capitalization; book-to-market ratio; stock turnover ratio; monthly stock volatility; firm age (number of months since IPO); stock beta; SP_500 dummy (equal to 1 if the stock belongs to the SP500 index); past 3-month returns; past-12-month returns. Except for beta, the SP_500 dummy, and the past-return measures, all variables are in natural logarithms. All returns are in percent. We report the time-series average of coefficient estimates from quarterly cross-sectional regressions, as well as the T-statistics (Newey-West adjusted standard errors).

Table 2 reports the regression results. The main message is that the duration measure based on passive funds holdings strongly predicts future stock returns.[18] For 3-month-ahead return regressions, the average coefficient on the ownership-weighted passive duration measure (equal-weighted passive duration measure) is 0.805 (0.968), statistically significant at the 1% level. The effect is also economically significant. For example, a one-standard-deviation increase in ownership- (equal-) weighted passive duration is associated with an increase in the 3-month-ahead returns of 0.48% (0.52%). The return predictability of the index fund holdings duration measure extends up to 2 years. The average coefficient for the ownership- (equal-) weighted passive duration is 3.133 (3.279) for the 1-year-ahead return regressions, and that for the 2-year-ahead return regressions is 2.665 (3.295) for the ownership- (equal-) weighted passive duration. All

[18] We also conduct regression analysis with index% only (1) and with index% and log(duration) together (2). Under both circumstances, index% is not statistically significant. Our results stress the role of long-term index holdings rather than index holdings at a certain period.

estimates are statistically significant at the 1% level. In terms of economic significance, a one-standard-deviation increase in ownership- (equal-) weighted passive duration is associated with a 1.89% (1.76%) increase in 1-year-ahead returns, and the corresponding increase is 1.61% (1.77%) for 2-year-ahead returns. Other control variables in our sample period are almost all statistically insignificant. In the appendix, we also find that the duration measure based on passive funds holdings can predict future excess returns.

Our second measure for passive funds' long-term investment in stocks is the churn ratio based on the quarterly holdings of passive funds, as defined previously in Section 3. Similar to the duration measure, we construct both equal- and ownership-weighted churn ratio measures. We repeat the same cross-sectional regression analysis, by replacing duration measures with the churn ratio. Table 3 presents the results. Column (1) reports the results for the ownership-weighted churn ratio measure, and Column (2) the equal-weighted measure. The results are consistent with those of the duration measures. Specifically, for the ownership-weighted measure, the average coefficient on the churn ratio is -2.056 for 3-month-ahead and statistically significant at the 1% level; and -6.940 and -9.016 for 1-year-ahead, and 2-year-ahead returns, respectively, and statistically significant at the 5% level. As the churn ratio is negatively correlated with the duration measure, the negative coefficient is consistent with the notion that stocks that are held longer by passive funds have higher future returns. In terms of economic significance, a one-standard-deviation increase in the churn ratio measure is associated with a 0.236% (0.798%, 1.037%) increase in 3-month (1-year, 2-year) ahead returns. Results for the equal-weighted churn ratio are negative but less significant. Overall, results based on churn ratio measures are consistent with those based on holdings duration measures. To save space, we will only present the duration-based results from now on. Churn ratio-based results are available upon request.

4.4.2. Passive holdings duration and future stock returns: A portfolio approach

At the end of each quarter, we rank stocks in our sample based on their weighted passive duration measure and form five portfolios, with portfolio 1 being the quintile with the shortest duration measure and portfolio 5 the longest. These portfolios are held for 3 months, 1 year, or 2 years, respectively. As a result, for 1-year and 2-year holding periods, there will be overlapping portfolios each quarter, similar to the design of the momentum portfolio strategy adopted by Jegadeesh and Titman (1993). For each month, monthly equal-weighted returns are recorded for each portfolio, as well as the average return of the overlapping portfolio returns for each quintile. Table 4 reports the time-series averages of monthly returns for each portfolio, as well as the return differences between the longest and shortest passive duration portfolios. For each holding horizon, we also report portfolio alphas from the time-series regressions of the four-factor model (the Fama and French (1993) three factors plus the momentum factor), and the five-factor model (the four factors plus the Stambaugh and Pastor (2003) liquidity factor). The results suggest that portfolios that have the longest passive duration outperform those with the shortest passive funds duration, consistent with the cross-sectional regression analysis. Specifically, when the holding period is 3 months, the monthly return difference between quintile 5 and quintile 1 is 0.57% based on portfolio raw returns. The corresponding difference in the four-factor (five-factor) alpha is 0.709% (0.705%) per month, and all differences are statistically significant at the 1% level. The results for portfolios with 1-year and 2-year holding periods are very similar. The performance difference is due to both the outperformance of the long holdings' duration portfolios and the underperformance of the short duration portfolios. For example, for the 3-month holding period, quintile portfolio 5 has a monthly 5-factor alpha of 0.423%, while the corresponding 5-factor alpha for quintile portfolio 1 is -0.282%, with both alphas being significant at the 5% level. We also rank stocks in our sample based on their weighted passive churn ratio measure and form five portfolios, with portfolio 1 being the quintile with the shortest churn ratio measure (longest holding) and portfolio 5 the longest churn ratio measure (shortest holding). We have similar results in Appendix A2. Overall, the results reported in Table 2 to Table 4 are consistent with hypothesis H1.

4.4.3 Return predictability of passive holdings duration: Evidence based on past stock performance

The evidence presented in the previous sections shows the strong return predictability of passive duration measures, which is consistent with passive funds' effective monitoring role in firms' management. If the permanent ownership of these passive funds provides incentives for them to closely monitor firms' governance and performance, their incentives should be especially strong for underperforming firms in which passive funds have had substantial holdings over a long period of time. The return predictability of passive holdings duration would in turn be stronger for those firms. We test this hypothesis (H2 in Section 2) next.

To examine this hypothesis, each quarter we split our sample into halves based on either the past 1-year or 3-year stock performance. We then create a dummy variable *low*, which equals one if the past 1-year (3-year) return is below the cross-sectional median, and 0 otherwise. We then include both the dummy variable *low* and the interaction term *log (dur-weighted)* **low* in the cross-sectional regressions. The interaction term captures the marginal return predictability of passive holdings duration for firms with low past returns. If passive funds have stronger incentive to monitor underperforming firms, we would expect the coefficient for the interaction term to be positive and significant. Table 5 reports the results. Columns on the left show results for the case in which the prior stock return performance is measured over the past 1-year period, and columns on the right show results where the past return performance is defined over the past 3-year period.

Consistent with the monitoring hypothesis, the predictability of the passive duration measure is stronger for firms with poor past 1-year (3-year) performance. For example, for the 3-month return regressions, the average coefficient on the passive duration measure is 0.321, which is only statistically significant at the 10% level. The interaction term involving the dummy variable *low*, however, is 0.981, which is significant at the 1% level. The implied coefficient on the passive holdings' duration for firms with low prior returns is 1.302, and is again significant at the 1% level. For 1-year-ahead return regressions, the average coefficient for the passive duration measure is 1.889, which is statistically significant at the 1% level. The coefficient for the interaction term is 2.481, again highly significant at the 1% level. This implies that for 1-year-ahead returns, the return predictability of passive holdings duration for firms with low past returns is more than twice as strong as that for high past return firms. For the regression specification involving 2-year-ahead returns, the coefficient on the interaction term becomes insignificant. The marginal effect of passive holdings duration seems to be diminished at the 2-year horizon[19].

Interestingly, although the past 1-year return is itself not a significant predictor of future returns in our sample, the low past return dummy variable strongly predicts future stock returns. For example, when the *low* dummy is defined over the past 1-year returns, the coefficient on the *low* dummy is -2.905 for 3-month-ahead returns, and statistically significant at the 1% level. This implies that on average, the next quarter return would be 0.85% lower for stocks in the bottom half of past 1-year returns, compared to stocks in the upper half. The results are qualitatively consistent for 1-year- and 2-year- ahead returns, and when *low* dummy is defined over the past 3-year returns.

[19] We also conduct analysis by splitting our sample into quintiles based on past 1-3 year returns and then introducing a quintile variable [-2, -1, 0, 1, 2] that we interact with duration. Our results remain the same. In the following subsample analysis, we also use quintile sorts as a robustness check but our results remain the same.

Overall, the evidence presented in Table 5 is consistent with hypothesis H2, in that passive funds have stronger incentive to monitor underperforming firms' management, and their monitoring is quite effective in bringing about performance improvements.

4.4.4 Return predictability of passive funds' holdings duration: Firm size and market conditions

We further hypothesize that monitoring would be more effective for smaller stocks and stocks with a greater degree of uncertainty, where passive funds can exert greater influence on the firms' management. We examine this hypothesis (H3 in Section 2) in this subsection.

Similar to the analysis based on past stock returns, each quarter we split our sample firms into two halves based on their market capitalization. We then create a dummy variable *small*, which equals one if the market capitalization is below the sample median value for the quarter, and zero otherwise. We include the *small* dummy and the interaction term *log (dur-weighted)* **small* in our regression specifications and examine the marginal effect of duration on smaller firms.

Panel A of Table 6 presents the results. For stocks at the bottom of the market capitalization, passive duration significantly predicts future stocks returns. For 3-month-ahead returns, the average coefficient for the duration measure is 0.282 and is statistically insignificant. The interaction term with *small* is, however, significant at the 5% level, with an average coefficient of 0.715. The implied coefficient on the passive duration measure for smaller stocks is 0.997 and significant at the 1% level. For 1-year-ahead returns, the average coefficient for passive duration is larger at 1.544, and only statistically significant at the 10% level. However, the coefficient for the interaction term is even larger at 2.231. The implied coefficient on the passive duration is 3.775

for smaller stocks, which is more than double that for larger stocks. For 2-year-ahead returns, the coefficient on the duration is insignificant again. The coefficient for the interaction term is less significant, with an average coefficient of 2.129, and the implied coefficient on the passive duration for smaller stocks is 3.199. Interestingly, in our sample, stocks in the smaller half of the market cap have lower returns on average. The marginal effect of the *small* stock dummy variable is reflected in its coefficient of -1.996 for 3-month-ahead returns, which implies that on average, the returns of stocks in the bottom half of the market capitalization are 0.51% lower for the next 3 months. Results are similar for 1-year- and 2-year-ahead returns.

In terms of market conditions, we expect that the passive funds' monitoring incentive would be stronger during more volatile periods, when there is greater uncertainty about the performance of the stocks they invest in. We use the CBOE VIX index as a proxy for market volatility, and repeat the cross-sectional regression analysis for the low market volatility periods (i.e., periods in which the VIX measure is below the sample median of 17%) and high market volatility periods (with VIX above the sample median value) separately. Panel B of Table 6 shows that the average coefficient for passive holdings duration during the high-volatility periods is 0.931 for 3-month-ahead returns. The coefficient for the 1-year- and 2-year-ahead returns is 4.111 and 3.509, respectively, and both coefficients are statistically significant at the 1% level. In contrast, holdings duration coefficients for the low-volatility periods remain less statistically significant. Their magnitudes are much smaller: The corresponding coefficients for 3-month-, 1-year-, and 2-year-ahead return regressions are 0.671, 2.159, and 1.674, respectively.

In summary, the results in Table 6 suggest that passive funds' monitoring is indeed more effective for small stocks and stocks with a greater degree of uncertainty, consistent with hypothesis H3.

4.4.5 Passive duration return predictability: Limited resources

Even though passive funds have strong incentives to monitor a firm's governance and improve stock performance, given their holdings of hundreds of stocks it is unlikely that they have the resources to pay equal attention to all of their stocks. We conjecture that the passive funds duration measure would have stronger predictability for stocks that are more important in passive funds' holdings (hypothesis H4 in Section 2).

4.4.5.1 A stock's importance in passive funds' holdings

To measure a stock's relative importance in passive funds' holdings, we calculate a stock's excess portfolio weight in passive funds. As discussed in Section 3, for each stock, we calculate the excess weight as the ratio of passive funds' dollar holdings of the stock relative to total net assets of passive funds, and then subtract the stock's percentage weight in the market portfolio (we use the U.S. domestic equity market as the proxy for the market portfolio.). Everything else equal, a stock would be relatively more important to passive funds if the excess weight is higher. We then split the sample into equal halves each quarter using the median value of the excess weight measure as the cutoff point. We define a dummy variable *important* that equals one if the stock is in the top half based on the relative weight measure, and zero otherwise. We then include the dummy variable and the interaction term with duration in the cross-sectional regressions. If stocks in the top half are indeed more important for passive funds in their holdings, we would expect passive funds to have stronger incentives and to allocate more resources for effective monitoring, and hence a positive coefficient on the interaction term.

Table 7 presents the regression results. Indeed, the interaction between passive holdings duration and the dummy variable *important* have a positive effect on future returns. The average coefficient on the interaction term is 0.576 (1.167) for the 3-month- (1-year-) ahead return regression, and is statistically significant at the 5% level. However, the average coefficient for the 2-year-ahead return regression is insignificant from zero. The passive duration itself is significant at the 5% level. Overall, the results presented in Table 7 show that the return predictability of the passive duration is much stronger for stocks that are more important in passive funds' overall portfolio holdings, consistent with hypothesis H4.

4.4.5.2 Russell 1000 vs Russell 2000 stocks

So far, our analysis is based on the entire sample of U.S. equity index funds and ETFs. Although we controlled for a number of firm characteristics in the cross-sectional regressions, it is quite possible that our specifications omit certain relevant variables. An alternative approach is to compare subgroups of stocks that otherwise have similar characteristics, but have different weights in passive funds' portfolio holdings. To do so, we follow Appel et al. (2016) to directly compare stocks at the bottom of the Russell 1000 and the top of the Russell 2000. Stocks near the cutoff boundaries of the indexes should share similar characteristics, including market capitalization. As the top 250 stocks in the Russell 2000 index have greater proportional weights in the index, however, the passive funds' ownership of these stocks would be much larger than that of stocks among the bottom 250 of the Russell 1000 stocks. We would then expect the predictability of the passive funds' holdings duration to be stronger for the top 250 stocks in the Russell 2000 compared to the bottom 250 stocks in the Russell 1000 index.

We require that the stocks in both the Russell 1000 and Russell 2000 indexes stay in the index for at least 2 consecutive years, and the stocks must also be represented in the S12 fund holdings data.[20] We then select the bottom 250 stocks in the Russell 1000 index and the top 250 stocks in the Russell 2000 index at the end of June each year. Panel A of Table 8 reports summary statistics for these two samples over the period 2011:Q2 – 2015:Q3. On average, we have 210 stocks from the bottom 250 stocks of the Russell 1000 index, and 230 stocks from the top 250 stocks of the Russell 2000 index in our sample. As expected, the average market capitalization of the bottom Russell 1000 stocks is larger than that of the Russell 2000 stocks. The average market capitalization for the bottom Russell 1000 stocks in our sample is $2.88 billion, and that for the top Russell 2000 stocks is $2.49 billion. The passive funds' percentage holding of the top Russell 2000 stocks (11.2%) are considerably higher than those in the bottom Russel 1000 stocks (8.4%). The difference in passive funds' ownership is quite significant at the 1% level. The average passive funds' holdings duration for the top Russell 2000 stocks is also higher, at 10.30 quarters, compared to that for the bottom Russell 1000 stocks at 9.68 quarters.

Panel B of Table 8 reports the regression results. Indeed, the return predictability of passive funds' holdings duration is only significant for the top Russell 2000 stocks. For the top Russell 2000 stocks, the average coefficient on the passive holdings' duration measure is 1.583 for the 3-month-ahead return regressions, with a t-statistic of 1.850. For 1-year- and 2-year-ahead return regressions, the corresponding coefficients are 5.208 and 8.680, and the associated t-statistics are 2.90 and 2.26, respectively. On the other hand, none of the coefficients on the passive duration measure is significant for the bottom Russell 1000 stocks. Consistent with evidence from tests based on stock excess weight, results for the bottom (top) Russell 1000 (2000) stocks suggest that

[20] Our results are qualitatively unchanged, as we impose the requirement that stocks continuously remain in the Russell 1000/2000 index for the past 5 years.

passive funds would indeed spend more efforts and be more effective in monitoring the performance of stocks that are more important in their portfolio holdings.[21]

4.4.6. Alternative explanations

One possible alternative explanation for the predictive ability of passive funds' holdings duration is that the increasing popularity of index funds and the investor flows to these funds lead to higher valuations for stocks in the relevant indexes. Figure 1 shows that the size of index funds has increased dramatically over our sample period. However, several aspects of the evidence from our empirical tests suggest that investor flow-driven price changes are unlikely to explain our findings. First, we require that all stocks have 2 consecutive quarters of passive funds' holdings data to be included in the sample. Therefore, short-term positive shocks to investor flows cannot directly explain the predictability of future returns at horizons of up to 2 years.

Second, it is well documented that investors chase past performance. Hence, investor flow-driven return predictability should be more pronounced for stocks that have been performing well. Our evidence, however, shows that the return predictability of passive holdings duration is stronger for poorly performing stocks. To more directly control the effect of fund flows, we include an additional control variable, namely, the percentage change in quarterly passive funds' holdings. If fund flows affect our results, we would expect the corresponding coefficient to be positive and significant, and after controlling for the fund flow effect, the impact of the passive holdings' duration on future stock returns should be weakened. However, we find that the coefficient for the percentage change in quarterly passive funds' holdings is insignificant. Controlling for the change

[21] We advise caution in interpreting the results presented in this table, as they are based on only 5 years of Russell 1000/2000 index constituent data we were able to obtain from FTSE Russell.

in passive fund holdings, the average coefficient for passive fund holdings duration remains virtually unchanged.

Another potential explanation for our findings is that the better performing stocks would mechanically enjoy a longer duration of ownership by passive funds' holdings. To the extent that such stocks continue to perform well in the future, one would expect a positive relation between passive funds' holdings duration and future stock returns. To explore this potential reverse causality, we examine the correlation between the duration of holdings measure and past stock returns. In untabulated results, we find that the correlations are in fact quite weak and, in some cases, negative.

4.5. Long-term return predictability: Passive funds vs active funds

We have presented evidence that duration measures based on passive fund holdings predict future stock returns, and our results are consistent with passive funds' incentive to monitor firms' governance and performance. A number of recent papers on mutual funds (e.g., Cremers and Pareek, 2016; Lan, Moneta, and Wermers, 2016) argue that long-term funds' investors may have information about firms' long-term performance, and their patient investment strategy outperforms passive funds, especially for active funds whose holdings are different from their benchmarks. Given the long-term nature of investments in both cases, it is possible that long-term active funds are also simply investing in stocks that benefit from passive fund ownership (Hypothesis H5).

4.5.1. Passive fund duration, active fund duration, closet index duration, and stock returns

We test Hypothesis H5 using the same Fama-Macbeth regression framework as in Section 4, by regressing future 3-month/12-month stock returns on duration measures based on active funds holdings and passive funds holdings. First, we decompose active funds into pure active funds and closet indexers. Following Cremers and Pareek (2016), we define those funds with more than 60% active shares as pure active funds, and the rest as closet indexers. Since the holdings' duration measure connects funds' overall holding history, we require that a fund has high/low active shares during the entire sample period. Next, we construct duration measures based on holdings of closet indexers and pure active funds, separately. Finally, for each return horizon (3-month and 1-year), we examine two models. In Model (1), we compare the effect of duration measures based on closet indexers and on active funds, and in Model (2) we also include the duration measure based on passive funds holdings.

Columns (1) and (3) of Table 9 report the results for Model (1). For 3-month returns, the coefficient of duration based on pure active funds is marginally significant at the 10% level, while it becomes insignificant for 12-month returns. The duration measure based on closet indexers is 0.071 at the 3-month horizon, albeit statistically insignificant. For the 1-year horizon, the coefficient is 0.872 and statistically significant at the 1% level. These results are consistent in spirit with Harford, Kecskes, and Mansi (2018) that holdings of long-term investors predict higher future returns.

Harford, et al. (2018) include closet indexers as index funds. Closet indexers might be able to mimic index funds, but they still have the flexibility to exit their stock positions. Hence, they do not necessarily have the same incentives as genuine index fund investors. We next test the marginal effects of duration measures based on passive funds holdings and closet indexer holdings, respectively. When we include the passive funds' holdings duration measure in Columns (2) and (4), the coefficients on active funds duration decline to 0.247 (0.492) for 3-month (12-month) returns, compared to Model (1), and they are statistically insignificant at the 10% level. The coefficient on closet indexers' duration declines to -0.013 (0.512) for 3-month (12-month) returns. It is statistically insignificant for 3-month returns, and only significant at the 10% level for 12-month returns. By contrast, the coefficient for passive funds holdings duration is 0.390 (3-months) and 1.948 (12-months), and remains statistically significant at the 5% level.[22]

4.5.2. Double sorts by stocks in passive funds duration and active funds duration

To further compare the long-term holdings' effect on stock returns between passive funds and active funds, we employ double sorts of stocks into 5x5 portfolios by passive fund duration and by active fund duration. Table 10 reports the results. In panel A, we first sort the sample of stocks into quintiles by the passive duration (*dur-weighted*) each quarter. Within each passive duration quintile, we further sort stocks into quintiles by the active fund duration (*dur-weighted-ac*) each quarter. In Panel B, we switch the order of sorting. Therefore, panel A examines the effect of active fund duration, controlling for passive fund duration, while Panel B examines the effect of passive fund duration, controlling for active fund duration.

These portfolios are held for 1 quarter and1 year, respectively. We calculate monthly equal-weighted returns for each portfolio. For each holding horizon, we also report portfolio alphas from the time-series regressions of the five-factor model (the Fama and French (1993) three factors plus the Carhart momentum factor and the Stambaugh and Pastor (2003) liquidity factor), as well as the difference in alphas between the longest and shortest active durations in panel A (passive durations in panel B).

[22] We also compare duration in passive funds and in active funds directly, without dividing active funds by active share. Results are similar: for 1-quarter returns and 1-year returns, after controlling for passive fund duration, the coefficient on active fund duration decreases in magnitude and is no longer statistically significant.

In panel A, for the 1-quarter holding period, the difference in monthly alphas is significant at the 5% level only for the two shortest passive fund duration quintiles, with the point estimates being 0.537%, and 0.343%, respectively. Alpha differences for the other three passive fund duration-sorted quintiles are statistically insignificant at the 10% level. Results for the 1-year holding period are similar. The alpha difference (0.556% per month) is significant at the 5% level only for the quintile with the shortest passive fund duration (the 1st quintile). For the next two passive duration quintiles (2nd, 3rd), alpha differences are smaller at 0.265% and 0.279% per month, and significant only at the 10% level. For the 4th and 5th passive fund duration quintiles, the differences in alpha are statistically insignificant.

In panel B, for each of the active fund duration-sorted quintile portfolios, alpha differences between the portfolio with the longest passive duration and that with the shortest passive duration are statistically significant for almost all subportfolios. For example, for portfolios with a 1-year holding period, the difference in the 5-factor alpha is 0.665% and statistically significant at the 1% level for the shortest active-fund duration quintile, and remains significant for quintiles with longer active fund duration. For quintile 5 with the longest active fund duration, the monthly alpha difference is 0.436% and significant at the 5% level. The results are similar for portfolios with a 1-quarter holding period. The only exception is for quintile 5 with the longest active fund duration, where the alpha difference becomes statistically insignificant at the 10% level.

Overall, our results show that it is the long-term ownership by the genuine index funds and ETFs, rather than the active funds' long-term investment that best predicts future stock returns. The evidence is consistent with our hypothesis H5, as described in Section 2.

4.6. Concluding Remarks

Recent years have witnessed a significant shift in investor interest from actively managed funds to low-cost passive funds designed to match the performance of market indexes. The implication of this shift for the governance of publicly traded firms owned by passive funds has been the subject of considerable interest and debate. The conventional view is that ownership by passive funds weakens corporate oversight. However, recent research on this issue has offered a very different viewpoints on this issue. Specifically, Appel, et al. (2016) demonstrate that passive fund investors do in fact play an important role in bringing about positive changes in firms' governance policies that lead to improvements in profitability and firm valuation. Motivated by these results, in this paper we further explore the implications of stock ownership by index funds for firms' stock performance over the short-term and the long-term.

We document a strong positive relation between the duration of passive fund holdings and subsequent performance of the stocks they own, both in the short term and at longer horizons of up to 2 years. The positive relationship between holdings duration and future stock returns is stronger in the case of poorly performing firms, smaller firms, and firms with larger proportional ownership by passive funds. Further, we find that the predictive ability of the passive funds' holdings duration measure for future stock returns is much stronger for stocks at the top of the Russell 2000 index compared to those at the bottom of the Russell 1000 index. These findings are consistent with the notion that significant holdings of passive funds are associated with more effective monitoring by the funds. We rule out a number of alternative explanations for our findings, including investor fund flow-driven price pressure and the potential for reverse causality. We also provide evidence that our results are not driven by closet indexers. Overall, the evidence in this study confirms that passive fund investors contribute to shareholder value creation. Since ‘exit’ is not an option for passive funds, they appear to bring about improvements in firm performance by actively engaging with the firms they own and exercising the power of their ‘voice’ over the long-term.

References

Appel, I.R., Gormley, T.A., Keim, D.B., 2016. Passive investors, not passive owners. Journal of Financial Economics, 121(1), pp.111-141.

Boone, A.L. and White, J.T., 2015. The effect of institutional ownership on firm transparency and information production. Journal of Financial Economics, 117(3), pp.508-533.

Carleton, W.T., Nelson, J.M. and Weisbach, M.S., 1998. The influence of institutions on corporate governance through private negotiations: Evidence from TIAA-CREF. Journal of Finance, 53(4), pp.1335-1362.

Cella, C., Ellul, A. and Giannetti, M., 2013. Investors' horizons and the amplification of market shocks. Review of Financial Studies, 26(7), pp.1607-1648.

Chakrabarty, B., Moulton, P.C. and Trzcinka, C., 2017. The performance of short-term institutional trades. Journal of Financial and Quantitative Analysis, 52(4), pp.1403-1428.

Cremers, M., Ferreira, M.A., Matos, P. and Starks, L., 2016. Indexing and active fund management: International evidence. Journal of Financial Economics, 120(3), pp.539-560.

Cremers, M. and Pareek, A., 2016. Patient capital outperformance: The investment skill of high active share managers who trade infrequently. Journal of Financial Economics, 122(2), pp.288-306.

Cremers, M., Pareek, A. and Sautner, Z., 2017. Short-term investors, long-term investments, and firm value. Unpublished working paper. University of Notre Dame.

Del Guercio, D. and Hawkins, J., 1999. The motivation and impact of pension fund activism. Journal of Financial Economics, 52(3), pp.293-340.

Fich, E.M., Harford, J. and Tran, A.L., 2015. Motivated monitors: The importance of institutional investors′ portfolio weights. Journal of Financial Economics, 118(1), pp.21-48.

Gaspar, J.M., Massa, M. and Matos, P., 2005. Shareholder investment horizons and the market for corporate control. Journal of Financial Economics, 76(1), pp.135-165.

Gillan, S.L. and Starks, L.T., 2000. Corporate governance proposals and shareholder activism: The role of institutional investors. Journal of Financial Economics, 57(2), pp.275-305.

Harford, J., Kecskes, A. and Mansi, S., 2018. Do long-term investors improve corporate decision making? Journal of Corporate Finance, 50, pp.424-452.

Jegadeesh, N. and Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. Journal of Finance, 48(1), pp.65-91.

Lan, C., Moneta, F. and Wermers, R., 2016. Holding horizon: A new measure of active investment management. Unpublished working paper. UNSW Business School.

Pástor, Ľ. and Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. Journal of Political Economy, 111(3), pp.642-685.

Smith, M.P., 1996. Shareholder activism by institutional investors: Evidence from CalPERS. Journal of Finance, 51(1), pp.227-252.

Wahal, S., 1996. Pension fund activism and firm performance. Journal of Financial and Quantitative Analysis, 31(1), pp.1-23.

Yan, X. and Zhang, Z., 2009. Institutional investors and equity returns: Are short-term institutions better informed? Review of Financial Studies, 22(2), pp.893-924.

Table 1. Summary statistics

Table 1 reports summary statistics. The sample consists of U.S. common stocks from 2003.q1 to 2015.q3. *Dur-weighted* is the weighted average time that a stock has been held by passive funds over the previous 5 years (in quarters), using passive ownership as a weight. *Dur-equal* is the equal-weighted average time a stock has been held by passive funds over the previous 5 years (in quarters). *CR-weighted* is the weighted average of turnover of passive funds holding a stock, using passive ownership as a weight. *CR-equal* is the equal-weighted average of turnover of passive funds holding a stock. *Index%* is percentage of shares held by passive funds. Duration, Churn Ratio, and ownership in active funds are similarly defined. These variables and other control variables are defined in Section 3. We eliminate stocks with missing market capitalization or book value of equity data, and stocks with prices below $1 or above $1,000. We require that a stock be held by one fund for at least two consequent quarters. Duration is winsorized at 1st percentiles, and expressed in quarters. Churn Ratio, ownership, and book-to-market ratio are winsorized at 1st and 99th percentiles.

Panel A: Time-series statistics of cross-sectional averages

Variable	Mean	Std dev.	Min	Median	Max
Passive Funds:					
Dur-weighted	7.910	1.458	5.287	7.581	10.179
Dur-equal	7.118	1.260	5.071	6.827	9.100
CR-weighted	0.118	0.027	0.084	0.116	0.213
CR-equal	0.132	0.019	0.104	0.127	0.181
Index%	0.047	0.019	0.010	0.047	0.079
Active Funds:					
Dur-weighted-ac	5.785	1.171	3.944	5.567	7.582
Dur-equal-ac	5.289	1.346	3.265	4.847	7.472
CR-weighted-ac	0.161	0.017	0.133	0.160	0.198
CR-equal-ac	0.182	0.023	0.142	0.182	0.229
Active%	0.103	0.007	0.084	0.105	0.114
Control Variables:					
# Stocks	3726	280	2595	3714	4303
Price	27.157	5.245	15.841	26.715	38.014
Size(millions)	4436.570	1130.440	2634.710	4194.000	6967.550
Btm	1.107	0.310	0.578	1.136	1.753
Volatility	0.121	0.025	0.094	0.112	0.180
Turnover	0.171	0.025	0.128	0.169	0.234
Age (months)	220.723	17.719	193.064	217.063	254.184
Beta	1.025	0.113	0.746	1.022	1.243
Ret(t,t-3)	0.038	0.109	-0.277	0.034	0.342
Ret(t-3,t-12)	0.133	0.236	-0.371	0.115	0.795
Ret(t-12,t-36)	0.374	0.410	-0.456	0.330	1.611

Panel B: Time-series mean of cross-sectional correlations

	Churn ratio_fund	Dur-weighted	Dur-equal	CR-weighted	CR-equal	Index%
Duration _fund	-0.625					
Dur-weighted		1.000				
Dur-equal		0.842	1.000			
CR-weighted		-0.434	-0.412	1.000		
CR-equal		-0.327	-0.380	0.728	1.000	
Index%		0.253	0.219	0.008	0.089	1.000

	Price	Size	Btm	Volatility	Turnover	Age	Beta	SP_500 dummy	Ret(t,t-3)	Ret(t-3,t-12)	Ret(t-12,t-36)
Dur-weighted	0.071	0.057	0.045	-0.161	-0.105	0.455	-0.062	0.238	0.014	-0.011	-0.080
Dur-equal	0.010	-0.009	0.094	-0.138	-0.144	0.448	-0.086	0.212	0.009	-0.037	-0.147
CR-weighted	0.016	0.092	-0.048	0.141	0.227	-0.208	0.182	-0.135	0.034	0.077	0.014
CR-equal	0.177	0.262	-0.093	0.037	0.270	-0.110	0.157	0.036	0.075	0.156	0.072
Index%	0.234	0.307	-0.070	-0.091	0.344	0.304	0.290	0.205	0.020	0.021	0.003

Table 2: Stock duration and future returns

This table reports quarterly Fama-Macbeth regression estimates for 1-quarter-ahead, 1-year-ahead, and 2-year-ahead stock returns on passive funds' stock duration and stock characteristics. Stock-level passive duration is *dur-weighted* by ownership weighted in Column (1) and *dur-equal* by equal weighted in Column (2) across all the passive funds holding that stock. Sample period is from 2003.q1 to 2015.q3. All variables except beta, SP500 index membership, and returns are expressed in natural logarithms. Returns are in percent. Standard errors are based on the Newey-West (1987) estimator. *, **, and *** denote significance at the 10%, 5%, and 1% level.

	Ret(t,t+3)		**Ret(t,t+12)**		**Ret(t+12, t+24)**	
	(1)	(2)	(1)	(2)	(1)	(2)
Intercept	1.590	1.392	2.937	2.760	1.007	0.483
	(0.853)	(0.754)	(0.297)	(0.278)	(0.104)	(0.049)
Log(Dur-weighted)	0.805***	0.968***	3.133***	3.279***	2.665***	3.295***
	(3.663)	(3.890)	(3.056)	(3.492)	(3.431)	(3.462)
Log(Size)	-0.144	-0.136	-0.146	-0.145	0.082	0.089
	(-0.767)	(-0.738)	(-0.245)	(-0.244)	(0.138)	(0.150)
Log(Btm)	0.164	0.143	1.568	1.513	1.453	1.391
	(0.615)	(0.535)	(1.356)	(1.288)	(1.507)	(1.413)
Log(Turnover)	-0.007	0.008	-0.817	-0.781	-0.437	-0.407
	(-0.032)	(0.033)	(-1.339)	(-1.255)	(-0.715)	(-0.651)
Log(Volatility)	0.541	0.498	2.719	2.588	3.477	3.442
	(0.632)	(0.584)	(1.510)	(1.425)	(1.357)	(1.325)
Log(Age)	-0.133	-0.142	-0.336	-0.265	-0.249	-0.338
	(-1.055)	(-1.033)	(-0.910)	(-0.799)	(-0.490)	(-0.619)
Log(Price)	-0.002	-0.002	0.007	0.007	0.005	0.005
	(-0.637)	(-0.624)	(0.824)	(0.857)	(0.949)	(1.027)
Beta	-0.192	-0.165	-0.912	-0.813	-1.148	-1.083
	(-0.511)	(-0.441)	(-1.368)	(-1.173)	(-1.638)	(-1.535)
SP_500 dummy	0.175	0.168	1.180	1.204	1.292	1.313
	(0.390)	(0.382)	(0.684)	(0.676)	(0.719)	(0.716)
Ret(t,t-3)	-0.003	-0.003	-0.025	-0.025	-0.041	-0.041
	(-0.291)	(-0.292)	(-0.839)	(-0.825)	(-1.559)	(-1.545)
Ret(t-12,t-3)	-0.001	-0.000	-0.036	-0.035	-0.029	-0.027
	(-0.043)	(-0.022)	(-1.214)	(-1.190)	(-1.354)	(-1.311)
Adjusted R-square	0.056	0.056	0.051	0.050	0.035	0.035
Quarters	51	51	51	51	48	48
Obs	188807	188807	180296	180296	159702	159702

Table 3: Stock churn ratio and future returns

This table provides quarterly Fama-Macbeth regressions for future 1-quarter-ahead, 1-year-ahead, and 2-year-ahead returns on stock churn ratio in passive funds and stock characteristics. Stock level churn ratio is *CR-weighted* by ownership weighted in Column (1) and *CR-equal* by equal weighted in Column (2) across all the passive funds holding that stock. Sample period is from 2003.q1 to 2015.q3. All variables except Churn Ratio, beta, SP500 index membership, and returns are expressed in natural logarithms. Returns are in percent. Standard errors are based on the Newey-West (1987) estimator. *, **, and *** denote significance at the 10%, 5%, and 1% level.

	Ret(t,t+3)		**Ret(t,t+12)**		**Ret(t+12,t+24)**	
	(1)	(2)	(1)	(2)	(1)	(2)
Intercept	2.370	2.195	6.213	5.851	4.197	3.741
	(1.261)	(1.169)	(0.598)	(0.568)	(0.423)	(0.381)
CR-weighted	-2.056***	-0.886	-6.940**	-3.512	-9.016**	-6.929
	(-3.445)	(-1.271)	(-2.280)	(-0.963)	(-2.626)	(-1.531)
Log(Size)	-0.143	-0.152	-0.190	-0.227	0.080	0.086
	(-0.755)	(-0.797)	(-0.328)	(-0.398)	(0.137)	(0.147)
Log(Btm)	0.176	0.169	1.578	1.542	1.426	1.376
	(0.664)	(0.642)	(1.382)	(1.338)	(1.481)	(1.427)
Log(Turnover)	-0.001	-0.008	-0.806	-0.825	-0.411	-0.405
	(-0.003)	(-0.036)	(-1.332)	(-1.347)	(-0.661)	(-0.618)
Log(Volatility)	0.527	0.501	2.636	2.549	3.431	3.326
	(0.614)	(0.581)	(1.454)	(1.399)	(1.347)	(1.304)
Log(Age)	0.075	0.112	0.475**	0.576***	0.356	0.452
	(0.642)	(0.937)	(2.294)	(2.752)	(0.909)	(1.199)
Log(Price)	-0.002	-0.002	0.007	0.007	0.003	0.004
	(-0.663)	(-0.604)	(0.717)	(0.783)	(0.616)	(0.727)
Beta	-0.170	-0.170	-0.782	-0.754	-1.048	-1.000
	(-0.454)	(-0.455)	(-1.165)	(-1.122)	(-1.480)	(-1.412)
SP_500 dummy	0.178	0.239	1.299	1.445	1.327	1.437
	(0.395)	(0.524)	(0.759)	(0.835)	(0.740)	(0.789)
Ret(t,t-3)	-0.003	-0.003	-0.023	-0.023	-0.039	-0.038
	(-0.239)	(-0.249)	(-0.749)	(-0.749)	(-1.532)	(-1.519)
Ret(t-12,t-3)	-0.000	-0.001	-0.036	-0.036	-0.028	-0.029
	(-0.036)	(-0.044)	(-1.201)	(-1.223)	(-1.360)	(-1.379)
Adjusted R-square	0.055	0.055	0.049	0.049	0.034	0.034
Quarters	51	51	51	51	48	48
Obs	188807	188807	180296	180296	159702	159702

Table 4: Portfolio approach

This table reports monthly equal-weighted portfolio raw returns and alphas after controlling for Fama French three factors (market factor, size factor, value factor), Carhart momentum factor and market liquidity factor (Pastor and Stambaugh,2003). Stocks are divided into quintiles each quarter from 2003.q1 to 2015.q3 according to stock duration in passive funds *dur-weighted*, with quintiles 1 and 5 consisting of short- and long-duration stocks, respectively. We then report returns for these five portfolios and the return differences, which are calculated over the next one quarter, next 1 year, and next 2 years. For returns longer than one quarter, we use the Jegadeesh and Titman (1993) approach to adjust overlaps. All reported returns are in percent per month. *, **, *** represent significance for return difference at 10%, 5%, and 1% levels. Standard errors are based on the Newey-West (1987) estimator. To save space, we only report ownership-weighted duration.

	Dur-weighted					
	Monthly Equal-Weighted Return and Alpha					
	1	2	3	4	5	5-1
Ret(t,t+3)						
Raw return	0.540	0.736	0.931	0.918	1.114	0.574***
	(0.943)	(1.324)	(1.798)	(1.862)	(2.252)	(3.281)
4-factor Alpha	-0.286	-0.074	0.120	0.129	0.422	0.709***
	(-1.816)	(-0.710)	(1.742)	(1.721)	(2.996)	(4.804)
5-factor Alpha	-0.282	-0.073	0.119	0.128	0.423	0.705***
	(-1.799)	(-0.690)	(1.726)	(1.693)	(3.025)	(4.757)
Ret(t,t+12)						
Raw return	0.656	0.843	0.963	1.007	1.145	0.489***
	(1.160)	(1.539)	(1.940)	(2.143)	(2.418)	(3.048)
4-factor Alpha	-0.211	-0.015	0.111	0.181	0.416	0.627***
	(-1.386)	(-0.149)	(1.848)	(2.779)	(3.324)	(4.592)
5-factor Alpha	-0.209	-0.013	0.111	0.179	0.416	0.625***
	(-1.353)	(-0.135)	(1.838)	(2.699)	(3.338)	(4.467)
Ret(t,t+24)						
Raw return	0.738	0.885	1.017	1.096	1.184	0.447***
	(1.590)	(1.976)	(2.412)	(2.742)	(2.926)	(3.349)
4- factor Alpha	-0.180	-0.021	0.119	0.218	0.403	0.584***
	(-1.210)	(-0.196)	(1.950)	(3.320)	(3.401)	(4.213)
5-factor Alpha	-0.182	-0.022	0.118	0.220	0.403	0.585***
	(-1.198)	(-0.201)	(1.915)	(3.366)	(3.385)	(4.126)

Table 5: Subsample results based on past stock returns

This table provides quarterly Fama-Macbeth regressions for future one-quarter-ahead, 1-year-ahead, and 2-year-ahead returns on stock duration in passive funds interacted with "*low*" dummy variable and stock characteristics. Each quarter, we divide the total sample by past 1- year (3-year) cumulative returns into halves. If past 1-year (3-year) returns are below the cross-sectional median, then *low* equals to one, else zero. Sample period is from 2003.q1 to 2015.q3. All variables except beta, SP500 index membership and returns are expressed in natural logarithms. Returns are in percent. Standard errors are based on the Newey-west (1987) estimator. *, **, and *** represent significance at the 10%, 5%, and 1% levels. To save space, we only report results based on ownership-weighted holdings duration measure.

	Ret(t,t+3)	**Ret(t,t+12)**	**Ret(t+12,t+24)**	**Ret(t,t+3)**	**Ret(t,t+12)**	**Ret(t+12,t+24)**
	Past 1-year cumulative returns			Past 3-year cumulative returns		
Intercept	4.254*	9.620	5.155	3.504*	7.820	3.780
	(1.847)	(0.870)	(0.534)	(1.801)	(0.908)	(0.437)
Log(Dur-weighted)	0.321*	1.889***	2.366***	0.568***	2.094**	1.997**
	(1.719)	(2.882)	(3.442)	(3.023)	(2.070)	(2.253)
Log(Dur-weighted)*low	0.981***	2.481***	0.531	0.576**	2.318***	1.595
	(3.922)	(3.233)	(1.287)	(2.337)	(3.414)	(1.391)
Low	-2.905***	-8.404***	-3.101***	-1.872***	-7.472***	-3.803
	(-5.924)	(-4.350)	(-4.015)	(-3.537)	(-4.673)	(-1.640)
Log(Size)	-0.046	0.163	0.377	-0.026	0.195	0.293
	(-0.353)	(0.325)	(0.676)	(-0.211)	(0.434)	(0.536)
Log(Btm)	0.119	1.429	1.372	0.153	1.535	1.359
	(0.461)	(1.238)	(1.406)	(0.578)	(1.218)	(1.398)
Log(Turnover)	0.082	-0.579	-0.309	0.094	-0.607	-0.035
	(0.302)	(-0.818)	(-0.433)	(0.338)	(-0.752)	(-0.043)
Log(Volatility)	0.236	2.071	2.718	0.182	2.295	2.614
	(0.365)	(1.219)	(1.215)	(0.295)	(1.432)	(1.299)
Log(Age)	-0.102	-0.217	-0.191	-0.019	0.013	0.115
	(-0.979)	(-0.652)	(-0.418)	(-0.199)	(0.041)	(0.294)
Log(Price)	-0.483	-1.169	-1.107	-0.617	-1.615	-1.192
	(-0.929)	(-0.982)	(-0.909)	(-1.195)	(-1.477)	(-1.037)

Beta	-0.261	-1.174*	-1.060	-0.308	-0.979	-1.050*
	(-0.716)	(-1.685)	(-1.596)	(-0.807)	(-1.392)	(-1.678)
SP_500 dummy	0.052	0.828	0.988	0.083	1.031	0.706
	(0.139)	(0.477)	(0.558)	(0.237)	(0.624)	(0.391)
Ret(t,t-3)	-0.010	-0.047*	-0.048**	-0.008	-0.034	-0.030
	(-0.956)	(-1.795)	(-2.354)	(-0.769)	(-1.289)	(-1.531)
Ret(t-12,t-3)	-0.003	-0.049*	-0.033**	0.001	-0.036	-0.016
	(-0.316)	(-1.909)	(-2.242)	(0.106)	(-1.342)	(-1.009)
Ret(t-36,t-12)				-0.001	-0.004	-0.003
				(-0.334)	(-0.701)	(-1.035)
Adjusted R-square	0.060	0.056	0.037	0.063	0.058	0.038
Quarters	51	51	48	51	51	48
Obs	188807	180296	159702	174820	167066	148537

Table 6: Subsample results by firm size and market conditions

This table reports estimates from the quarterly Fama-Macbeth regressions for future one-quarter-ahead, 1-year-ahead, and 2-year-ahead returns on stock duration in passive funds and stock characteristics divided by firm size and market conditions. In Panel A, we divide the total sample by firm market capitalizations into halves each quarter. If the stock size is lower than cross-sectional median, then *small* dummy equals to one, else zero. In Panel B, we divide sample periods into halves by CBOE *VIX* index median (17%) in our sample period. Sample period is from 2003.q1 to 2015.q3. All variables except beta, SP500 index membership, and returns are expressed in natural logarithms. Returns are in percent. Standard errors are based on the Newey-West (1987) estimator. *, **, *** represent significance at 10%, 5%, and 1% confidence intervals. To save space, we only report ownership-weighted duration.

Panel A: Subsample results by firm size

	Ret(t,t+3)	**Ret(t,t+12)**	**Ret(t+12,t+24)**
Intercept	4.681**	14.604	10.589
	(2.198)	(1.362)	(1.349)
Log(Dur-weighted)	0.282	1.544*	1.070
	(1.021)	(1.683)	(0.777)
Log(Dur-weighted)*small	0.715**	2.231***	2.129
	(2.081)	(2.964)	(1.344)
Small	-1.996**	-7.709***	-6.369
	(-2.463)	(-4.504)	(-1.578)
Log(Size)	-0.163	-0.592	-0.212
	(-1.276)	(-1.107)	(-0.567)
Log(Btm)	0.126	1.408	1.324
	(0.520)	(1.259)	(1.389)
Log(Turnover)	0.039	-0.833	-0.490
	(0.145)	(-1.189)	(-0.702)
Log(Volatility)	0.240	2.091	2.841
	(0.364)	(1.208)	(1.275)
Log(Age)	-0.082	-0.206	-0.122
	(-0.798)	(-0.639)	(-0.276)
Log(Price)	-0.448	-1.063	-1.080
	(-0.870)	(-0.901)	(-0.848)
Beta	-0.225	-1.026	-0.976
	(-0.555)	(-1.392)	(-1.425)
SP_500 dummy	0.277	2.059	2.137
	(0.714)	(1.213)	(1.524)
Ret(t,t-3)	-0.002	-0.019	-0.032
	(-0.227)	(-0.725)	(-1.664)
Ret(t-12,t-3)	0.002	-0.027	-0.022
	(0.212)	(-1.047)	(-1.370)
Adjusted R-square	0.061	0.055	0.038
Quarters	51	51	48
Obs	188807	180296	159702

Panel B: Subsample results based on market conditions

	Ret(t,t+3)		Ret(t,t+12)		Ret(t+12,t+24)	
	VIX>=17%	VIX<17%	VIX>=17%	VIX<17%	VIX>=17%	VIX<17%
Intercept	3.797	1.378	10.566	-0.829	12.292**	-6.752
	(1.072)	(0.655)	(1.228)	(-0.096)	(2.599)	(-0.543)
Log(Dur-weighted)	0.931***	0.671**	4.111***	2.159**	3.509***	1.674**
	(3.624)	(2.221)	(5.305)	(2.477)	(6.420)	(2.804)
Log(Size)	0.048	-0.109	0.227	0.208	0.724	0.128
	(0.267)	(-0.640)	(0.464)	(0.367)	(0.845)	(0.233)
Log(Btm)	-0.174	0.439	0.922	2.025	1.433	1.428
	(-0.525)	(1.309)	(0.973)	(1.435)	(1.104)	(1.069)
Log(Turnover)	0.525	-0.381***	0.491	-1.749***	0.447	-1.144***
	(1.098)	(-2.850)	(0.376)	(-7.191)	(0.346)	(-4.491)
Log(Volatility)	-0.205	0.656	0.647	3.326	6.857**	-1.261
	(-0.189)	(0.883)	(0.459)	(0.895)	(2.484)	(-0.320)
Log(Age)	-0.040	-0.156	-0.268	-0.185	-1.280***	0.932***
	(-0.216)	(-1.614)	(-0.899)	(-0.770)	(-4.270)	(3.639)
Log(Price)	-1.509	0.596*	-3.781**	1.648***	-3.239**	1.209
	(-1.710)	(1.965)	(-2.452)	(3.164)	(-2.305)	(1.596)
Beta	-0.220	-0.289	-1.258	-1.065	-2.031	-0.128
	(-0.307)	(-1.126)	(-1.338)	(-1.250)	(-1.117)	(-0.197)
SP_500 dummy	-0.553	0.546	-0.462	1.882	2.009	-0.227
	(-1.178)	(1.245)	(-0.410)	(1.273)	(0.833)	(-0.253)
Ret(t,t-3)	-0.014	0.008	-0.064	0.023*	-0.084**	0.021
	(-0.968)	(0.603)	(-1.242)	(1.757)	(-2.664)	(1.465)
Ret(t-12,t-3)	-0.005	0.010*	-0.058	0.002	-0.067***	0.024***
	(-0.260)	(2.035)	(-1.214)	(0.272)	(-5.475)	(3.110)
Adjusted R-square	0.080	0.040	0.065	0.044	0.037	0.037
Quarters	25	26	25	26	24	24
Obs	92215	96592	88393	91903	80148	79554

Table 7: Importance of underlying stocks holding in passive funds

This table provides quarterly Fama-Macbeth regressions for future one-quarter-ahead, 1-year-ahead, and 2-year-ahead returns on stock duration in passive funds interacted with "*Important*" dummy and stock characteristics. Stock excess weight is measured as the difference between a stock's passive holding weight in total passive funds and the stock's value weight in market portfolios. We then sort the total sample by the excess weight into halves each quarter. If a stock's excess weight is above the cross-sectional median, *important* equals to one, else zero. Sample period is from 2003.q1 to 2015.q3. Standard errors are based on the Newey-West (1987) estimator. *, **, *** represent significance at the 10%, 5%, and 1% levels. To save space, we only report ownership-weighted duration.

	Ret(t,t+3)	**Ret(t,t+12)**	**Ret(t+12,t+24)**
Intercept	3.230	5.186	1.004
	(1.609)	(0.525)	(0.112)
Log(Dur-weighted)	0.564**	2.540***	2.517**
	(2.564)	(2.856)	(2.575)
Log(Dur-weighted)*important	0.576**	1.167**	-0.182
	(2.140)	(2.610)	(-0.159)
Important	-1.431**	-1.500	2.568
	(-2.116)	(-1.080)	(0.914)
Log(Size)	-0.078	0.329	0.769
	(-0.632)	(0.845)	(1.569)
Log(Btm)	0.139	1.438	1.382
	(0.619)	(1.290)	(1.491)
Log(Turnover)	0.087	-0.699	-0.467
	(0.330)	(-0.908)	(-0.617)
Log(Volatility)	0.227	2.064	2.866
	(0.340)	(1.250)	(1.299)
Log(Age)	-0.060	-0.306	-0.408
	(-0.636)	(-0.811)	(-0.888)
Log(Price)	-0.446	-1.036	-1.048
	(-0.895)	(-0.881)	(-0.842)
Beta	-0.268	-1.236*	-1.142*
	(-0.552)	(-1.737)	(-1.753)
SP_500 dummy	0.105	0.811	0.740
	(0.305)	(0.487)	(0.436)
Ret(t,t-3)	-0.003	-0.019	-0.029
	(-0.262)	(-0.720)	(-1.528)
Ret(t-12,t-3)	0.002	-0.026	-0.021
	(0.254)	(-1.026)	(-1.319)
Adjusted R-square	0.060	0.055	0.038
Quarters	51	51	48
Obs	188807	180296	159702

Table 8: Stocks in Russell 1000 Index vs Russell 2000 Index

This table compares stocks in Russell 1000 and Russell 2000 indexes. We select the sample as: (1) a stock is held by Russell 1000 (Russell 2000) at the end of June in the previous year and (2) this stock is ranked in the bottom 250 of Russell 1000 (top 250 of Russell 2000) at the end of June in this year. Panel A compares the summary statistics between the two groups. We provide the mean level of each variable, the difference of the mean between the two groups and associated t-values after clustering on individual firms. Panel B provides quarterly Fama-Macbeth regressions for future one-quarter-ahead, 1-year-ahead, and 2-year-ahead returns on stock duration and stock characteristics by comparing stocks in Russell 1000 and Russell 2000 indexes. Sample period is from 2011.q2 to 2015.q3. Standard errors are based on the Newey-West (1987) estimator. *, **, *** represent significance at the 10%, 5%, and 1% levels. To save space, we only report ownership-weighted duration.

Panel A: Summary statistics

	Bottom 250 stocks of Russell 1000	Top 250 stocks of Russell 2000	Difference	T statistics
	(1)	(2)	(1)-(2)	(1)-(2)
Size (1000s)	2878.956	2489.414	389.542	2.65
Dur-weighted	9.684	10.301	-0.617	-3.85
CR-weighted	0.090	0.081	0.009	3.24
Index%	0.084	0.112	-0.028	-11.52

Panel B: Fama-Macbeth regressions

	Ret(t,t+3)		**Ret(t,t+12)**		**Ret(t+12, t+24)**	
	Russell1000	Russell2000	Russell1000	Russell2000	Russell1000	Russell2000
Intercept	3.739	8.534	2.408	64.029	26.858	83.541**
	(0.446)	(0.596)	(0.100)	(1.439)	(1.493)	(2.276)
Log(Dur-weighted)	1.115	1.583*	-0.606	5.208***	-3.999	8.680**
	(1.033)	(1.850)	(-0.216)	(2.901)	(-0.839)	(2.260)
Log(Index%)	0.449	-0.503	3.785	-4.189	5.325*	-6.767***
	(0.740)	(-0.363)	(1.527)	(-1.101)	(1.845)	(-3.609)
Log(Size)	0.628	-0.632	4.645*	-4.592	-0.189	-8.387*
	(0.736)	(-0.526)	(1.824)	(-0.999)	(-0.143)	(-2.111)
Log(Btm)	-0.064	-0.488	-0.544	-0.247	-2.750**	0.328
	(-0.129)	(-0.742)	(-0.445)	(-0.131)	(-2.612)	(0.207)
Log(Turnover)	-0.48	-2.548***	-1.675	-7.824***	-3.108*	-3.814***
	(-0.798)	(-3.381)	(-0.800)	(-6.450)	(-1.906)	(-3.148)
Log(Volatility)	-1.797	1.798	-5.073	1.862	2.753	-2.039
	(-1.353)	(1.697)	(-1.707)	(0.603)	(0.603)	(-0.308)
Log(Age)	-1.068**	0.119	-3.233**	0.292	-0.833	0.31
	(-2.129)	(0.371)	(-2.352)	(0.371)	(-0.417)	(0.377)
Log(Price)	-0.004	-0.02	0.005	0	-0.01	0.079
	(-1.018)	(-1.183)	(0.364)	(0.001)	(-1.052)	(1.625)
Beta	0.855	0.064	-2.983**	-1.852	-9.276***	-1.114
	(0.887)	(0.082)	(-2.357)	(-0.719)	(-3.183)	(-0.398)
Ret(t,t-3)	-0.018	-0.006	-0.143*	0.058	-0.033	0.029
	(-0.693)	(-0.243)	(-1.922)	(0.820)	(-0.284)	(0.444)
Ret(t-12,t-3)	0.018	0.021	0.054	0.06	0.082	-0.013
	(1.001)	(1.551)	(1.254)	(1.631)	(1.340)	(-0.573)
Adjusted R-square	0.094	0.095	0.094	0.090	0.114	0.103
Quarters	18	18	18	18	15	15
Obs	3413	3831	3344	3728	2723	2989

Table 9: Comparing passive funds and active funds

This table provides quarterly Fama-Macbeth regressions for future one-quarter-ahead and 1-year-ahead on stock duration and stock characteristics. We first divide active mutual funds into closet indexers and pure active funds by active share (cutoff 60%) following Cremers and Pareek (2016), and only select the funds that continuously belong to either group during the sample period. In columns (1) and (3), we compare stock duration in closet indexers and in pure active funds. Next, in columns (2) and (4), we introduce stock duration in passive funds and compare the long-term holding effect of passive funds, closet indexers, and pure active funds respectively**.** Sample period is from 2003.q1 to 2015.q3. All variables except beta, SP500 index members, and returns are expressed in natural logarithms. Returns are in percent. Standard errors are based on the Newey-West (1987) estimator. *, **, *** represent significance at the 10%, 5%, and 1% levels. To save space, we only report ownership-weighted duration.

	Ret(t,t+3)		**Ret(t, t+12)**	
	(1)	(2)	(3)	(4)
Intercept	4.163*	3.728*	10.308	8.229
.	(1.771)	(1.767)	(0.912)	(0.741)
Log(Dur-weighted)		0.390**		1.948**
		(1.979)		(2.579)
Log(Dur-weighted-closet indexers)	0.071	-0.013	0.872***	0.512*
	(0.520)	(-0.096)	(3.068)	(1.797)
Log(Dur-weighted-active funds)	0.308*	0.247	0.759	0.492
	(1.842)	(1.605)	(0.956)	(0.688)
Log(Size)	-0.117	-0.109	-0.206	-0.148
.	(-0.790)	(-0.710)	(-0.343)	(-0.246)
Log(Btm)	-0.132	-0.138	0.177	0.179
.	(-0.454)	(-0.543)	(0.151)	(0.154)
Log(Turnover)	0.011	0.003	-0.138	-0.172
.	(0.035)	(0.011)	(-0.163)	(-0.206)
Log(Volatility)	0.121	0.155	1.398	1.582
.	(0.183)	(0.223)	(0.829)	(0.949)
Log(Age)	-0.002	-0.066	-0.063	-0.404
.	(-0.012)	(-0.482)	(-0.167)	(-0.869)
Log(Price)	-0.499	-0.486	-1.101	-1.064
.	(-1.077)	(-1.105)	(-0.958)	(-0.939)
Beta	-0.319	-0.315	-1.293**	-1.330**
	(-0.588)	(-0.539)	(-2.076)	(-2.171)
SP_500 dummy	0.015	0.059	0.903	0.798
	(0.044)	(0.197)	(0.528)	(0.479)
Ret(t,t-3)	-0.009	-0.009	-0.016	-0.017
.	(-0.621)	(-0.676)	(-0.451)	(-0.477)
Ret(t-12,t-3)	0.000	0.000	-0.017	-0.016
	(0.038)	(0.037)	(-0.563)	(-0.553)
Adjusted R-square	0.070	0.071	0.064	0.065
Quarters	51	51	51	51
Obs	128234	128234	123455	123455

Table 10: Double sort by stock duration in passive funds and active funds

This table reports monthly equal-weighted double sort (5*5) portfolio five-factor alphas after controlling for Fama French three factors, Carhart momentum factor and Pastor and Stambaugh (2003) market liquidity factor. Panel A first sorts all the stocks each quarter into quintiles by passive fund ownership-weighted duration. Then within each quintile, stocks are second sorted into quintiles by active fund ownership-weighted duration. Panel B switches the sequence: First sort all the stocks into quintiles each quarter by active fund ownership-weighted duration, and second within each quintile group, sort stocks into quintiles by passive fund ownership-weighted duration. The portfolios are held for either 3 months or 12 months. We report the monthly five-factor alphas (in percent) as well as the difference in alphas between portfolio 5 and portfolio 1 for the second sorting sequence. We follow Jegadeesh and Titman (1993) to adjust overlaps. *, **, *** represent significance at 10%, 5%, and 1% level. Standard errors are based on the Newey-West (1987) estimator.

Panel A: First sort by passive duration (rows), and then sort by active duration (columns)

		Ret (t, t+3)						**Ret (t, t+12)**					
		Second sort: Stock duration in active funds (Dur-weighted-ac)						**Second sort: Stock duration in active funds (Dur-weighted-ac)**					
		1	2	3	4	5	5-1	1	2	3	4	5	5-1
1	5-factor alpha	-0.628	-0.383	-0.138	-0.171	-0.091	0.537**	-0.528	-0.341	-0.379	-0.051	0.028	0.556**
		(-3.177)	(-2.552)	(-0.909)	(-1.020)	(-0.465)	(2.223)	(-2.512)	(-1.861)	(-3.472)	(-0.319)	(0.118)	(2.246)
2	5-factor alpha	-0.141	-0.259	-0.073	-0.049	0.202	0.343**	-0.216	-0.130	-0.035	-0.000	0.049	0.265*
		(-1.108)	(-1.699)	(-0.608)	(-0.368)	(1.337)	(1.995)	(-1.569)	(-1.091)	(-0.272)	(-0.003)	(0.307)	(1.695)
3	5-factor alpha	-0.031	0.025	0.076	0.185	0.199	0.230	-0.109	-0.076	0.149	-0.001	0.170	0.279*
		(-0.258)	(0.281)	(0.995)	(2.191)	(1.642)	(1.334)	(-0.759)	(-0.857)	(1.338)	(-0.005)	(1.618)	(1.972)
4	5-factor alpha	-0.042	0.129	0.129	0.107	0.211	0.253	0.186	0.097	0.194	0.172	0.247	0.061
		(-0.354)	(1.414)	(1.615)	(1.331)	(1.893)	(1.612)	(1.832)	(1.743)	(3.807)	(1.711)	(2.207)	(0.462)
5	5-factor alpha	0.132	0.196	0.264	0.329	0.458	0.326	0.225	0.291	0.423	0.510	0.610	0.385
		(0.809)	(1.680)	(2.044)	(2.789)	(1.764)	(1.596)	(1.297)	(2.749)	(2.741)	(2.922)	(1.910)	(1.339)

Panel B: First sort by active duration (rows), and then sort by passive duration (columns)

		Ret (t, t+3)						Ret (t, t+12)					
		Second sort: Stock duration in passive fund(Dur-weighted)						Second sort: Stock duration in passive fund(Dur-weighted)					
		1	2	3	4	5	5-1	1	2	3	4	5	5-1
1	5-factor alpha	-0.649	-0.192	-0.305	-0.231	-0.168	0.481***	-0.579	-0.257	-0.345	-0.310	0.086	0.665***
		(-3.072)	(-1.343)	(-2.016)	(-1.923)	(-1.059)	(2.668)	(-2.528)	(-1.662)	(-2.326)	(-2.028)	(0.493)	(3.080)
2	5-factor alpha	-0.198	0.023	-0.017	0.159	0.198	0.396**	-0.014	-0.109	-0.173	-0.011	0.245	0.259*
		(-1.156)	(0.188)	(-0.191)	(1.972)	(1.421)	(2.240)	(-0.098)	(-1.024)	(-2.163)	(-0.106)	(2.216)	(1.898)
3	5-factor alpha	-0.076	0.017	0.064	-0.022	0.156	0.231*	0.042	0.085	-0.003	0.153	0.360	0.318**
		(-0.465)	(0.158)	(0.801)	(-0.292)	(1.448)	(1.685)	(0.303)	(0.745)	(-0.030)	(1.848)	(2.852)	(2.429)
4	5-factor alpha	0.024	0.004	0.187	0.278	0.309	0.286**	-0.003	0.058	0.108	0.156	0.398	0.401***
		(0.160)	(0.042)	(2.143)	(3.380)	(2.926)	(2.204)	(-0.024)	(0.578)	(1.499)	(2.393)	(2.753)	(3.461)
5	5-factor alpha	0.160	0.181	0.251	0.357	0.376	0.216	0.203	0.188	0.280	0.435	0.640	0.436**
		(1.173)	(1.738)	(2.214)	(3.992)	(1.610)	(1.034)	(1.309)	(1.665)	(2.731)	(4.327)	(2.462)	(2.022)

Figure 1: Time-series trends of fund ownership

This figure plots time-series trends of passive fund ownership and active fund ownership from 2003 to 2015

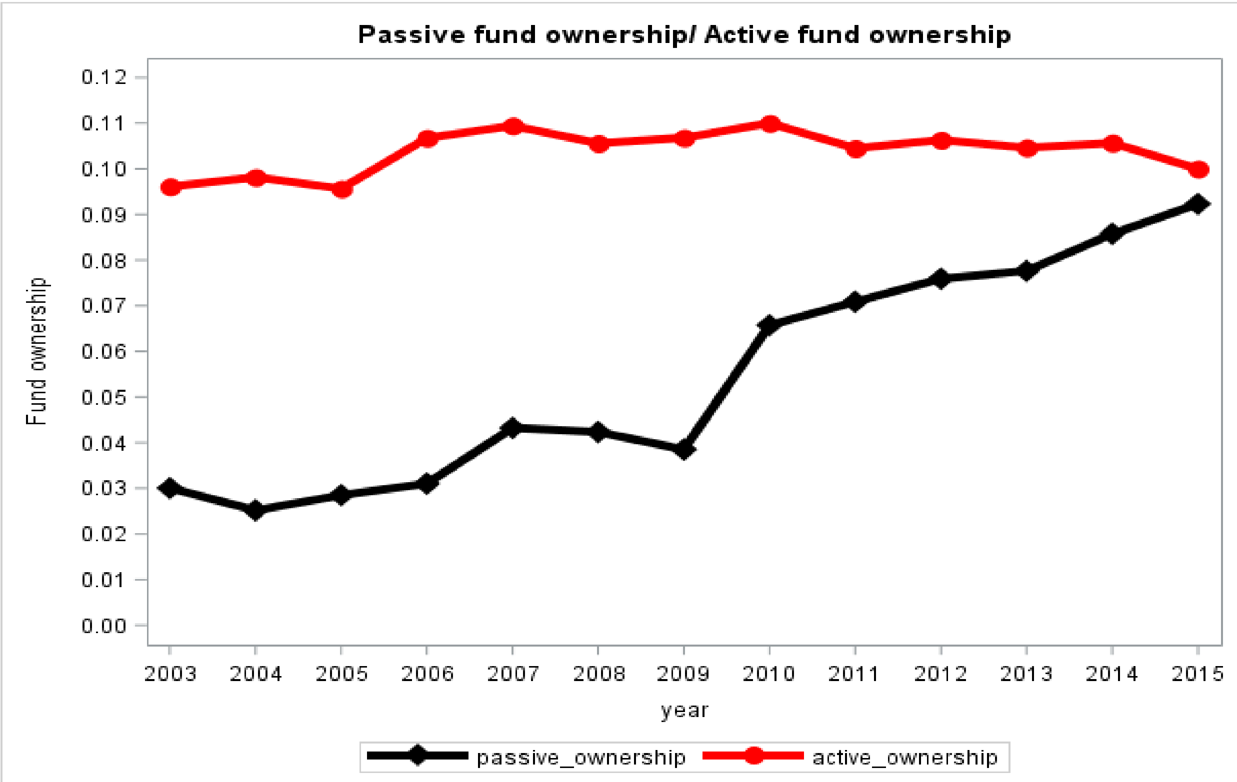


Figure 2: Time-series trends of stock level duration

This figure plots time-series trends of stock duration in passive funds and in active funds from 2003.q1 to 2015.q3

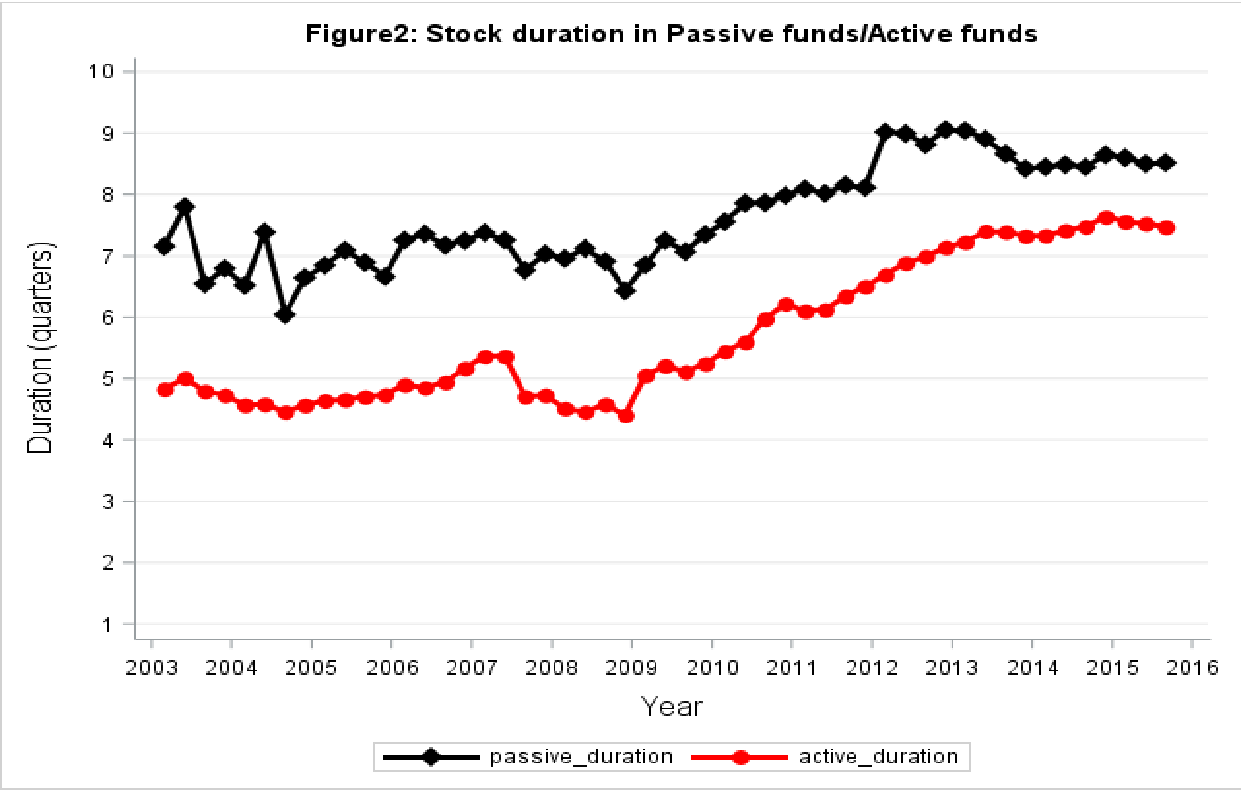
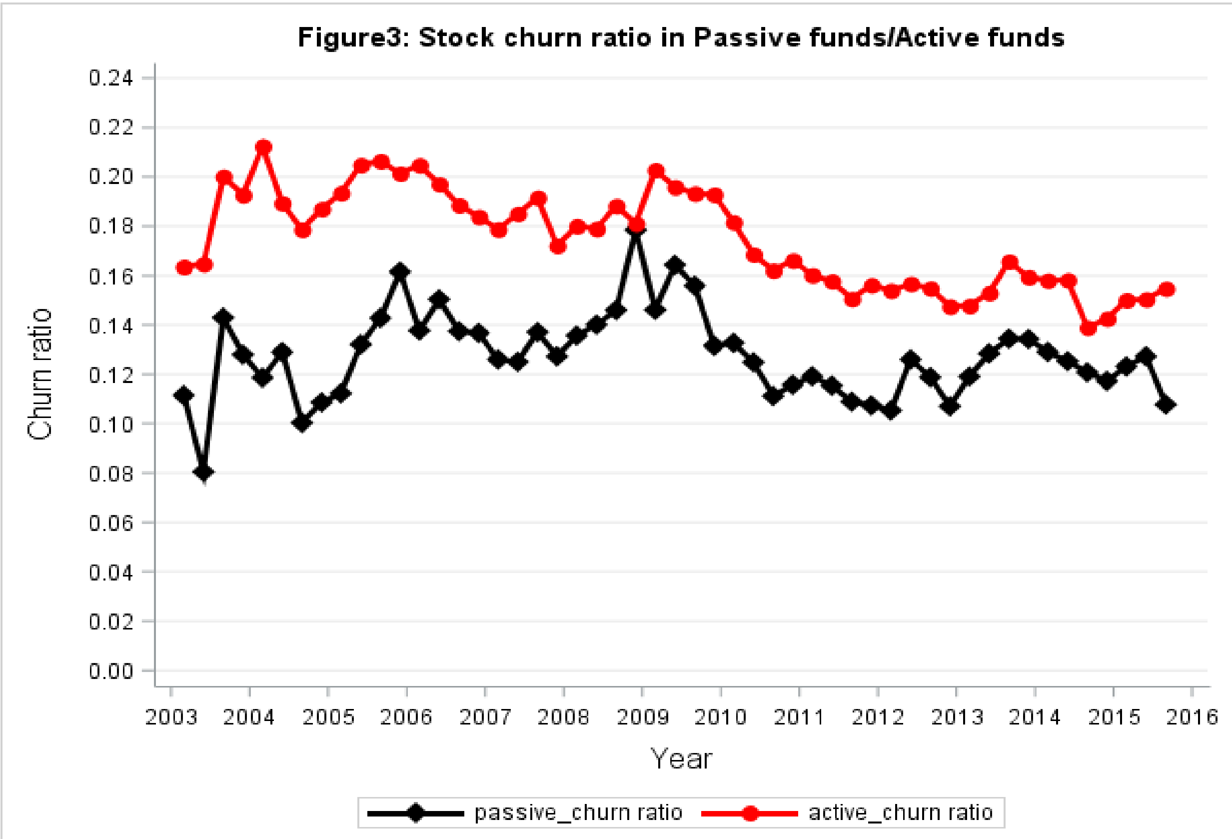


Figure 3: Time-series trends of stock level churn ratio

This figure plots time-series trends of stock churn ratio in passive funds and in active funds from 2003.q1 to 2015.q3



Appendix Table A_1: Stock duration and future excess stock returns

This table reports estimates from quarterly Fama-Macbeth regressions for future one-quarter-ahead, 1-year-ahead and 2-year-ahead returns on stock weighted-duration in passive funds and stock characteristics. Excess returns are calculated as raw returns minus risk-free rates in Column (1), raw returns minus value weighted market returns in Column (2), and raw returns minus value-weighted industry returns, which use Fama French 49 industry classifications, in Column (3). Sample period is from 2003.q1 to 2015.q3. All variables except beta, SP500 index membership, and returns are expressed in natural logarithms. Returns are in percent. Standard errors are based on the Newey-West (1987) estimator. *, **, *** represent significance at the 10%, 5%, and 1% level. To save space, we only report ownership-weighted duration.

	Ret(t,t+3)			**Ret(t,t+12)**			**Ret(t,t+24)**		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Intercept	2.256	-0.155	0.304	3.717	-6.933	-2.489	1.780	-7.769	-3.095
	(0.982)	(-0.079)	(0.198)	(0.348)	(-0.944)	(-0.536)	(0.173)	(-1.124)	(-0.737)
Log(Dur-weighted)	0.799***	0.783***	0.792***	3.110***	2.843***	2.583***	2.584***	2.481***	2.077***
	(3.850)	(3.856)	(4.478)	(3.186)	(3.453)	(4.661)	(3.478)	(3.642)	(4.191)
Log(Size)	-0.033	-0.026	0.002	0.205	0.300	0.261	0.414	0.472	0.328
	(-0.243)	(-0.196)	(0.016)	(0.406)	(0.631)	(0.803)	(0.718)	(0.862)	(0.751)
Log(Btm)	0.137	0.138	0.139	1.446	1.251	1.070	1.389	1.268	0.966
	(0.521)	(0.533)	(0.752)	(1.279)	(1.178)	(1.653)	(1.450)	(1.413)	(1.631)
Log(Turnover)	0.065	0.074	0.033	-0.629	-0.560	-0.728	-0.333	-0.247	-0.586
	(0.237)	(0.272)	(0.117)	(-0.895)	(-0.892)	(-1.206)	(-0.470)	(-0.374)	(-0.944)
Log(Volatility)	0.232	0.160	0.044	1.946	1.773	0.690	2.735	2.503	1.284
	(0.352)	(0.239)	(0.063)	(1.173)	(1.132)	(0.422)	(1.222)	(1.207)	(0.692)
Log(Age)	-0.099	-0.087	-0.186*	-0.233	-0.094	-0.533**	-0.183	-0.089	-0.443
	(-0.951)	(-0.837)	(-1.878)	(-0.717)	(-0.302)	(-2.098)	(-0.406)	(-0.213)	(-0.995)
Log(Price)	-0.436	-0.378	-0.363	-1.036	-0.694	-0.601	-1.037	-0.753	-0.395
	(-0.841)	(-0.769)	(-0.724)	(-0.893)	(-0.676)	(-0.581)	(-0.843)	(-0.678)	(-0.354)
Beta	-0.257	-0.246	-0.254	-1.155	-0.681	-0.249	-1.071	-0.723	-0.028
	(-0.700)	(-0.668)	(-0.799)	(-1.643)	(-1.001)	(-0.331)	(-1.621)	(-1.150)	(-0.053)
SP_500 dummy	0.008	-0.049	-0.111	0.749	0.119	-0.075	0.903	0.306	0.112
	(0.022)	(-0.127)	(-0.290)	(0.434)	(0.069)	(-0.046)	(0.498)	(0.174)	(0.067)
Ret(t,t-3)	-0.002	-0.002	-0.005	-0.020	-0.017	-0.025	-0.032*	-0.026	-0.024*
	(-0.215)	(-0.176)	(-0.633)	(-0.758)	(-0.713)	(-1.349)	(-1.698)	(-1.515)	(-1.803)

Ret(t-12,t-3)	0.002	0.003	-0.000	-0.027	-0.023	-0.024	-0.022	-0.015	-0.012
	(0.213)	(0.272)	(-0.004)	(-1.049)	(-1.025)	(-1.226)	(-1.372)	(-1.128)	(-1.254)
Adjusted R-square	0.060	0.059	0.050	0.054	0.053	0.048	0.037	0.036	0.032
Quarters	51	51	51	51	51	51	48	48	48
Obs	188807	188807	184540	180296	180296	176259	159702	159702	156912

Appendix A_2: Portfolio approach

This table reports monthly equal-weighted portfolio raw returns and alphas after controlling for Fama French three factors (market factor, size factor, value factor), Carhart momentum factor and market liquidity factor (Pastor and Stambaugh,2003). Stocks are divided into quintiles each quarter from 2003.q1 to 2015.q3 according to stock churn ratio in passive funds c*r-weighted*, with quintiles 1 and 5 consisting of short- and long-churn ratio stocks, respectively. We then report returns for these five portfolios and the return differences, which are calculated over the next one quarter, next 1 year, and next 2 years. For returns longer than one quarter, we use the Jegadeesh and Titman (1993) approach to adjust overlaps. All reported returns are in percent per month. *, **, *** represent significance for return difference at 10%, 5%, and 1% levels. Standard errors are based on the Newey-West (1987) estimator. To save space, we only report ownership-weighted duration.

CR-weighted						
Monthly Equal-Weighted Return and Alpha						
	1	2	3	4	5	5-1
Ret(t,t+3)						
Raw return	0.978	0.904	0.842	0.768	0.736	-0.242
	(1.918)	(1.911)	(1.664)	(1.377)	(1.252)	(-1.521)
4-factor Alpha	0.308	0.153	0.041	-0.071	-0.131	-0.439***
	(1.745)	(2.109)	(0.542)	(-0.786)	(-0.982)	(-2.979)
5-factor Alpha	0.312	0.154	0.039	-0.070	-0.130	-0.441***
	(1.788)	(2.128)	(0.520)	(-0.772)	(-0.968)	(-3.028)
Ret(t,t+12)						
Raw return	1.047	0.990	0.937	0.844	0.775	-0.272**
	(1.990)	(2.128)	(1.913)	(1.626)	(1.384)	(-2.105)
4-factor Alpha	0.332	0.193	0.095	-0.032	-0.130	-0.462***
	(1.742)	(3.480)	(2.048)	(-0.499)	(-1.032)	(-3.040)
5-factor Alpha	0.334	0.193	0.094	-0.031	-0.128	-0.462***
	(1.783)	(3.425)	(1.959)	(-0.484)	(-0.999)	(-3.081)
Ret(t,t+24)						
Raw return	1.137	1.094	0.990	0.862	0.822	-0.314**
	(2.473)	(2.738)	(2.420)	(2.012)	(1.832)	(-2.556)
4- factor Alpha	0.372	0.244	0.099	-0.066	-0.126	-0.499***
	(1.986)	(4.033)	(2.076)	(-0.926)	(-0.940)	(-2.907)
5-factor Alpha	0.371	0.244	0.100	-0.067	-0.128	-0.498***
	(1.973)	(4.029)	(2.073)	(-0.908)	(-0.922)	(-2.892)