UNIL | Université de Lausanne
Département de finance

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

FINANCE RESEARCH SEMINAR SUPPORTED BY UNIGESTION

"Innovation and informed trading: Evidence from industry ETFs"

Professor Maureen O'Hara
Cornell University

Friday, February 22, 2019, 10:30-12:00
Room 126, Extranef building at the University of Lausanne

UNIGESTION

Innovation and informed trading: Evidence from industry ETFs

Shiyang Huang, Maureen O'Hara, and Zhuo Zhong*

December 2018

We hypothesize that industry exchange traded funds (ETFs) encourage informed trading on underlying firms through facilitating hedging of industry-specific risks. We show that short interest on industry ETFs, reflecting part of the *"long-the-stock/short-the-ETF"* strategy, positively predicts returns on these ETFs and the percentage of positive earnings announcements of underlying stocks. We also show that hedge funds' long-short strategy using industry ETFs and industry ETF membership reduces post-earnings-announcement-drift. Our results suggest that financial innovations such as industry ETFs can be beneficial for informational efficiency by helping investors hedge risks.

JEL Classification: G12; G14
Keywords: ETF, hedge funds, short interest, market efficiency, financial innovation

Corresponding author: Maureen O'Hara, Johnson College of Business, Sage Hall, Cornell University, Ithaca, NY 14853, mo19@cornell.edu, 607 255 3645

---

* Shiyang Huang (huangsy@hku.hk) is at the Faculty of Business and Economics, the University of Hong Kong; Maureen O'Hara (maureen.ohara@cornell.edu) is at the Johnson College of Business, Cornell University and UTS; Zhuo Zhong (zhuo.zhong@unimelb.edu.au) is at the University of Melbourne. We thank Li An (discussant), Vikas Agarwal (discussant), Carole Comerton-Forde, Vyacheslav (Slava) Fos (discussant), Vincent Gregoire, Bruce Grundy, Terry Hendershott, Raymond Kan, Marta Khomyn (discussant), Bryan Lim, Andrea Lu, Dong Lou, Kalle Rinne (discussant), Valeri Sokolovski, Chen Yao (discussant), Liyan Yang as well as seminar participants at University of Columbia, University of Toronto, University of Melbourne, University of Technology Sydney, McGill, HEC Montreal, Wilfrid Laurier University, 2018 Luxembourg Asset Management Summit, 2018 UW Summer Finance Conference in Seattle, 2018 LSE Paul Woolley Center Conference, CICF 2018, Bank of America Merrill Lynch Asia Quant Conference 2018, Hong Kong-Shenzhen Greater Bay Area Finance Conference 2018, and FIRN 2018 for helpful comments and suggestions.

Innovation and informed trading: Evidence from industry ETFs

Few financial innovations in recent times have had the impact of exchange traded funds (ETFs). With assets approaching $3.5 trillion, ETFs are now larger than hedge funds. Worldwide there are approximately 5000 exchange traded funds, making ETFs the preferred investment approach for a wide range of investors. Indeed, it does not seem an exaggeration to argue that the growth of passive investing via ETFs has posed a "disruptive innovation" for the entire asset management industry.[1] For many investors, the main innovation of ETFs is to provide a more liquid, lower-cost alternative to mutual funds. For others, the innovation is access to previously unavailable asset classes. In this paper, we argue that another, perhaps under-appreciated, innovation is an expanded ability to hedge. We demonstrate that this aspect of ETF innovation has a direct impact on the nature of informed trading and the efficiency of the market.

We hypothesize that ETFs can facilitate hedging for informed investors. We develop this hypothesis based both on the theoretical literature on financial innovation and on industry reports on ETFs. The literature shows that financial innovations such as introducing a new security can improve risk-sharing (Allen and Gale, 1994). Moreover, Dow (1998) and Simsek (2013a, 2013b) show that the new security could enhance investors' arbitrage profits if it could be used to hedge their arbitrage risk. This idea of hedging is also widely observed in practice, especially in reports on how hedge funds use ETFs. For example, Bloomberg recently reported *"Hedge funds mainly use ETFs to take short positions. … As a group, hedge*

[1] Ananth Madhavan makes the case for such disruptive innovation in his book *Exchange Traded Funds and the New Dynamics of Investments*, (FMA Survey and Synthesis: 2016).

funds have $105 billion in short ETF positions — more than double their $43 billion in long positions. … The funds' shorts don't necessarily indicate bearish sentiment, but rather are used to hedge out part of the market in order to isolate a long position." [2]

To investigate this hedging role, we focus on the role played by industry ETFs. Because a stock's risk usually includes market risk, industry risk, and firm-specific risk, an informed trader hoping to profit from firm-specific information will want to hedge the market risk and the industry risk. While index futures or index ETFs are used to hedge the market risk, the advent of industry ETFs provides a vehicle to hedge the industry risk. The relatively low shorting cost of ETFs also makes hedging with industry ETFs more accessible. To our knowledge, ours is the first paper to address these industry-hedging effects.

We first establish two important facts. We show that the industry ETF is more likely to experience larger short interest than either non-industry ETFs or individual stocks. [3] Specifically, on average, the short interest ratio (SIR, short interest/shares outstanding) of the industry ETF is about 12%, while the SIR of the non-industry ETF is about 4%. Moreover, the industry ETF's SIR is highly skewed with a SIR of 60% at the 95 percentile. In contrast, the non-industry ETF (or the individual stock) has 17% at the same percentile, even though shares outstanding are comparable with the industry ETF's.[4]

Second, we find that large short interest on the industry ETF does not always imply a bearish outlook. Conditioning on the market state, we disentangle hedging-motivated and speculation-motivated short interest in industry ETFs. During the 2007-2008 crisis period, we find that speculation-motivated short sales mainly drive short interest, reflecting a bearish outlook on the industry. During the non-crisis period, however, hedging-motivated short

---

[2] See Bloomberg Intelligence, September 8th 2017.

[3] In Appendix (Figure A.1), we include a table from Bloomberg showing that industry ETFs are the most heavily shorted ETFs.

[4] Stock level short interest is also much lower than that of industry ETFs. The details are reported in the Appendix (Table A.2).

sales become the main driver. Specifically, during the non-crisis period, we find that large short interest predicts more positive earnings surprises among underlying stocks of the industry ETF.

This hedging-motivated short-selling has important implications for the return predictability of the industry ETF. Because a hedging-motivated short sale is not a directional bet on industry performance, sizable short interest on the industry ETF can generate a temporary price impact which will revert in the future. Consequently, short interest could positively predict the return of the industry ETF. Using a portfolio-sorting approach and Fama-MacBeth regressions, we confirm this prediction. We find that the change in short interest positively predicts the future return for an industry ETF, whereas we find the opposite pattern for short interest on constituent stocks of the industry ETF.[5] That is, at the member stock level, the change of stock-level short interest negatively predicts the future return of the stock, a result consistent with prior studies (e.g., Diether, Lee, and Werner, 2009; Rapach, Ringgenberg, and Zhou, 2016). Our finding that short interest on industry ETFs has the opposite implication from short interest on member stocks is a new, and we believe important, result.[6]

These findings are consistent with the hypothesis that informed investors use a *"long-the-stock/short-the-industry ETF"* strategy to hedge their long positions on stock with positive firm-specific information. To further support our hypothesis, we investigate the channel through which this trading occurs. We focus on hedge fund trading since hedge

[5] There are several reasons why we use the change of short interest as the sorting variable. First, as argued by the prior studies (e.g., Desai et al, 2002), the change of short interest could capture the information flows, which is more consistent with the theory of Diamond and Verrecchia (1987). Second, we find short interest on industry ETFs are persistently higher than non-industry ETFs. Third, short interest on industry ETFs is itself quite persistent (see Table 2). The last two points indicate using the change of short interest can avoid the potential ETF fixed effect.

[6] We also find that the short interest on industry ETFs negatively predict returns on these ETFs during the crisis period.

funds are likely to be trading on information and, unlike some other institutional investors (e.g., mutual funds), are not constrained in their ability to implement long-short strategies. Conditional on the stock having positive earnings surprise, we find hedge fund long positions on stocks are contemporaneously correlated with short interest on industry ETFs before earnings announcements. As placebo tests, we repeat the analysis using non-hedge funds (e.g., mutual funds and banks) and find no such contemporaneous correlations between these institutions' long position on stocks and short interest on industry ETFs. These results provide direct evidence linking the *"long-the-stock/short-the-ETF"* strategy to informed investors aiming to profit from positive firm-specific fundamentals. We further show that jointly considering abnormal hedge fund holdings and short interest on industry ETFs can positively predict stock returns, lending credence to hedge funds being the source of the informed trading.

If the ability to hedge with industry ETFs incentivizes informed investors to trade more aggressively on their information, then the market could become more informationally efficient. Focusing on earnings announcement events of industry ETFs' member stocks, we find that the *"long-the-stock/short-the-ETF"* strategy reduces the post-earnings-announcement-drift in response to the earnings surprise. Thus, the market is less surprised by positive earnings announcements, reflecting that information has already been at least partially impounded in prices through the informed investors' long-short strategy.

Although less pronounced, we also find evidence on the reverse *"long-the-ETF/short-the-stock"* strategy carried out by hedge funds betting on negative firm-specific fundamentals. This effect only holds for large stocks, however, consistent with the limiting effect imposed by the higher shorting costs of individual stocks. Overall, these results show that industry ETFs expand informed investors' ability to hedge and thereby improves market efficiency.[7]

Our paper contributes to the literature in several areas. The financial innovation literature shows that an important motive for creating a new financial security is to complete the financial market (see Duffie and Rahi (1995) for a comprehensive survey), enabling investors to better span their investment opportunities. Investors can isolate some risks with the new financial security which could lead to more optimal portfolios (Chen, 1995) and potentially more informed trading. Dow (1998) shows that informed investors with better hedging trade more aggressively on information. Simsek (2013a, 2013b) argues that financial innovation leads to more speculation among investors with different beliefs on different risk factors, as hedging becomes easier. Recently, a small theoretical literature studies the impact of ETFs. Cong and Xu (2016) show that introducing composite securities facilitates trading common factors in assets' liquidation values. Bhattacharya and O'Hara (2017) study how inter-market information linkages in ETFs can exacerbate market instability and herding. We provide direct empirical evidence showing that financial innovation, in particular, the industry ETF, is associated with more aggressive informed trading from sophisticated investors such as hedge funds.

Our paper also contributes to the growing empirical ETF literature. Past studies center around the price impact of arbitrage activity between the ETF and its underlying. Ben-David, Franzoni, and Moussawi (2018) find the ETF arbitrage activity increases non-fundamental volatility on underlying stocks. Wermers and Xue (2015) and Madhavan and Sobczyk (2015) also find ETFs are associated with higher volatility of the underlying. Da and Shive (2017)

[7] We are aware of the selection issue and thus use the propensity score matching to control firm characteristics (e.g., firm size and book-to-market ratio) to study the market efficiency implication of industry ETF memberships. In addition to post-earnings-announcement-drift, we also consider alternative measures of price efficiency, including price delay and variations ratio.

show the ETF arbitrage contributes to return co-movement. Baltussen, Bekkum, and Da (2018) show the ETF arbitrage also partially contributes to the change of index return serial dependence from positive to negative. While the literature seems to agree that ETFs increase volatility and co-movement in the underlying, the impact on liquidity and informational efficiency remains undetermined. Madhavan and Sobczyk (2015) find that ETFs have heterogeneous effects on price efficiency of underlying assets and the effect depends on the liquidity of ETFs. Glosten, Nallareddy, and Zou (2017) find ETF's membership positively affects informational efficiency at the stock level, especially, the incorporation of earnings information. Easley, Michayluk, O'Hara, and Putnins (2018) find that the market overall is not becoming more passive, and that informational efficiency is not being negatively affected. On the other hand, Israeli, Lee, and Sridharan (2017) find the ETF ownership is associated with a larger bid-ask spread and less informative price. Our paper provides a new perspective to study the ETFs' impact on informational efficiency. We focus on industry ETFs and test the channel that, by facilitating informed traders' hedging needs, industry ETFs encourage more informed trading.

Last but not least, our paper adds to the literature on short selling. Although past empirical findings largely imply short sellers have superior information in predicting future abnormal returns (see Desai et al., 2002; Asquith, Pathak, and Ritter, 2005; Boehmer, Jones, and Zhang, 2008; and Diether, Lee, and Werner, 2009), the literature also acknowledges some short sellers are merely hedgers (see Chen, Da, and Huang, 2018). Hedgers short a stock to hedge their positions in other assets, for example, hedging a convertible bond purchase or the delta hedge in put option trading. Ofek, Richardson, and Whitelaw (2004) show short-sale constraints in the underlying stock increase violations of put-call parity, suggesting that the difficulty in hedging an option position affects prices in both the stock and option market. Battalio and Schultz (2011) and Grundy, Lim, and Verwijmeren (2012) study the 2008 short-sale ban and find that option bid-ask spreads increase for banned stocks. Their results imply option markets are disrupted when hedging the underlying stock with short-sales becomes difficult (or almost impossible). Our analysis of short interest on industry ETFs provides additional evidence on the impact of hedging-based short-sales. Our results also show an intriguing asymmetry as the hedging effect is stronger for positive news, but weaker for negative news. We conjecture it is due to the higher short-sale costs of individual stocks, making a *"short-the-stock/long-the-ETF"* strategy less implementable.

1. Data description and sample statistics

Our study uses two sets of data. The first data set contains information on U.S. industry ETFs. For each U.S. industry ETF, we track its short interest, holdings, price, and volume from its inception to December 2017. The second data set contains the earnings announcements of all publicly listed firms from January 1999 to December 2017. We complement the above datasets with a variety of related data such as hedge fund holdings, non-hedge fund holdings, and firm characteristics. In this section, we discuss the construction of our two primary data sets in detail.

1.1 The ETF level data

A. The equity ETF

To construct the list of industry ETFs on U.S. equity, we first need the list of U.S. equity ETFs. We start with the fund universe of the CRSP Survivor-biased-free Mutual Fund database. We identify a fund as an ETF if the "et_flag" of the fund is "F." Also, we require these funds to have the CRSP share-code of either "44" or "73." To obtain the non-synthetic U.S. equity ETF, we drop funds whose name contains "bond," "bear," or "hedged." After those steps, we merge our list with a snapshot of all U.S. equity ETFs from ETFDB in June

2018.[8] For each ETF, we track its holdings information from the inception date to December 2017.[9] To ensure our list contains only equity ETFs, we apply a filter which requires our sample ETFs to have at least 80% investment in U.S. common domestic stocks. Our final sample consists of 508 U.S. equity ETFs, which is close to past studies.[10]

B. The industry ETF

We extract industry ETFs from the abovementioned equity ETFs based on holdings information. We match an ETF's holdings with the Fama-French 12 industry classification, and then identify the industry in which the ETF has the most investment. To qualify for an industry ETF, we require the dominating industry investment exceed one-third of the ETF's portfolio size. This requirement gives us 244 industry ETFs. We filter out ETFs whose name contains "value," "growth," "Russell," "dividend," "momentum," or "dynamic" to ensure the ETF is primarily aiming for a specific industry coverage. After this step, we are left with 144 ETFs. We further require that the ETF consists of at least 30 stocks (in the Appendix we remove this requirement and show that our results hold with a less restrictive industry ETF list). Finally, we obtain a list of 121 industry ETFs covering 10 out of 12 industries in the Fama-French classification. Figure 1 shows the time series growth of the total net asset value and the number of industry ETFs in our sample.[11]

[Insert Figure 1 Here]

---

[8] ETFDB is a website providing detail information on ETF, see www.etfdb.com for details.
[9] We use 13F data from Thompson Reuters for fund holdings, and complement it with the CRSP Survivor-biased-free Mutual Fund database.
[10] Glosten, Nallareddy, and Zou (2017) identify 447 ETFs between 2004 and 2013; Israeli, Lee, and Sridharan (2017) identify 443 ETFs between 2000 and 2014; Da and Shive (2017) identify 549 ETFs between 2006 and 2013; Li and Zhu (2017) identify 343 ETFs from 2002 to 2013.
[11] The earliest ETF in our sample starts in January 1993 and the earliest industry ETF in our sample starts in December 1998.

C. The price, volume, and short interest data for equity ETFs

We obtain the monthly price and volume for our ETF sample from CRSP. The monthly short interest for both equity ETFs and their underlying stocks are from COMPUSTAT. We collect all those data from the inception date to December 2017. We define the short interest ratio as short interest over total shares outstanding. Panel A and B in Table 1 report the summary statistics of short interest, total shares outstanding, price, volume, and the net asset value for our ETF sample. We report the industry and non-industry ETF, separately.

[Insert Table 1 Here]

1.2 The firm level data

A. Data on earnings announcements

We construct our data on earnings announcements based on analyst-target-price forecasts from the Institutional Brokers' Estimate System (I/B/E/S), quarterly financial statements from COMPUSTAT, and financial market data from CRSP. Our sample period is from January 1999 to December 2017. We focus on quarterly earnings announcements that are available in both COMPUSTAT and I/B/E/S.[12] Following Livnat and Mendenhall (2006) and other papers in this literature, we impose the following restrictions:

(1). Ordinary common shares listed on the NYSE, AMEX, or NASDAQ.

(2). The earnings announcement date is reported in both COMPUSTAT and I/B/E/S, and the earnings report dates in COMPUSTAT and in I/B/E/S differ by no more than one calendar day.

[12] We use the link table provided by Byoung-Hyoun Hwang from Cornell University. This link table provides a mapping from I/B/E/S ticker to CRSP permno and can be downloaded from his webpage: http://www.bhwang.com/code.html.

(3). The price-per-share at the end of the fiscal quarter is available from COMPUSTAT and is greater than $1.

(4). The market value of equity at the fiscal quarter-end is available and is larger than $5 million.

(5). Daily stock returns are available in CRSP for the dates around the earnings announcement. Moreover, the stock should be assignable to one of the six Fama-French benchmark portfolio based on size and book-to-market ratio.

We define an earnings surprise by the standardized unexpected earnings (SUE). The SUE of firm i at quarter t is calculated as $SUE_{i,t} = \frac{EPS_{i,t} - EPS_{i,t-4}}{\sigma_{i,t}}$, where EPS_t is the earnings per share at quarter t, and EPS_{t-4} is the earnings per share at the same quarter in the previous year. $\sigma_{i,t}$ is the standard deviation of $EPS_{i,t} - EPS_{i,t-4}$ in the last eight consecutive quarters.

B. The hedge fund and non-hedge fund list

We construct a list of hedge funds based on Form ADV (an SEC regulatory filing). After 2011, all U.S. hedge fund advisers with more than $150 million in asset under management are required to file Form ADV. Following Brunnermeier and Nagel (2005), Griffin and Xu (2009), and Jiang (2017), we take two steps to construct the hedge fund list. First, an asset management adviser from Form ADV is identified as a hedge fund if 80% of its assets are in the hedge fund business (as reported in Form ADV). Second, the list of hedge funds in the first step is manually merged with Form 13 (CDA/Spectrum) through advisers' names.[13] The CDA/Spectrum database is also used to construct the list of U.S. non-hedge funds. Following Lou (2012), mutual funds in our sample have a minimum fund size of $1 million, and the total net assets, TNA, reported by the CDA/Spectrum do not differ from CRSP's TNA by more than a factor of 2 (TNA from CDA/Spectrum should between one half and double of

[13] The detailed description can be referred to the online Appendix of Jiang (2017).

CRSP's TNA). The equity mutual funds in our sample have investment objective codes: aggressive growth, growth, growth and income, balanced, unclassified, or missing.

C. Data on institutional holdings

We follow Ben-David, Franzoni, and Moussawi (2011) to construct institutional holdings for each firm at each quarter based on the Thompson Reuters 13F data.[14] Merging this with the above list on hedge funds, we obtain hedge fund holdings on our sample firms at each quarter. To estimate abnormal holdings, we take the difference between the current quarter holdings and the moving average of the past four quarters holdings. Similarly, we obtain non-hedge fund abnormal holdings on our sample firms at each quarter. In Panel C Table 1, we report summary statistics for our earnings announcements sample after winsorizing at the bottom and top 1%. All variables have a distribution similar to past studies.[15]

2. Short interest on industry ETFs

Can industry ETFs facilitate informed trading and enhance the efficiency of the market? In this section, we approach this question by first examining the behavior of short interest on industry ETFs. We study whether short interest on industry ETFs reflects speculation or hedging. We then focus on a particular hedging strategy we call the *"long-the-stock/short-the-ETF"* strategy, and examine its relation to exploiting firm-specific information. To explore further the implication of hedging-motivated short interest on industry ETFs, we investigate the relation between short interest and future returns on industry ETFs, and the relation between ETF-level short interest and stock-level hedge funds trading performance.

2.1 Why do investors short industry ETFs?

[14] WRDS provides the detail code for constructing institutional holdings from the 13F data, see https://wrds-web.wharton.upenn.edu/wrds/research/applications/ownership/Institutional%20Trades/.

[15] Our hedge fund abnormal holdings have a similar magnitude on the mean and standard deviation as Chen, Da, and Huang (2018).

Panels A and B in Table 1 compare the short interest ratio (SIR) of industry and non-industry ETFs. We make two observations. First, on average, the SIR of the industry ETF is higher than non-industry ETFs (12% versus 4%). Second, the SIR of industry ETFs has a much longer right tail than that of non-industry ETFs. The industry ETF has a SIR of 60% at the 95th percentile whereas the latter has less than 20% at the same percentile.[16] Figure 2 shows the histogram of the SIR. For industry ETFs, we indeed observe a significant concentration of the SIR at the 100% level. Such a pattern is not observed among non-industry ETFs. The longer right tail of the SIR indicates that industry ETFs tend to experience more extreme short positions.[17] These observations are also consistent with the Bloomberg summary of most 15 shorted ETFs at the July 2017 (see Figure A.1)

[Insert Figure 2 Here]

Table 2 provides further evidence that industry ETFs are more likely to have extreme short interest. We sort all ETFs into quintiles based on the SIR. We find that industry ETFs are more likely to end up in the top SIR quintile, whereas non-industry ETFs are more likely to fall into the bottom SIR quintile. For example, within the top quintile of the SIR, 60% are industry ETFs; while the percentage of non-industry ETFs is about 84% within the bottom quintile of the SIR. Furthermore, industry ETFs that are in the high SIR quintile are more likely to remain in the high SIR quintile in the following month. The likelihood of staying is about 88% based on the Markov transition matrix estimate.

---

[16] The difference is not due industry ETFs having fewer shares outstanding as the distribution of the shares outstanding of the industry ETF is quite similar to that of the non-industry ETF.

[17] In constructing the short interest ratio, we replace all ratios above 100% with 100%. In other words, the concentration of the short interest ratio at 100% represents a large cumulative mass of short interest exceeding 100%. In Table A.2, we also find that ETF constitute stocks are less likely to have SIR in terms of mean and 95th percentile.

[Insert Table 2 Here]

What drives extreme short interest in industry ETFs? One natural explanation is that investors are betting against a specific industry, e.g., investors use financial-sector ETFs to short the financial industry during the 2008 financial crisis. We call this the speculation or speculative short sale hypothesis. An alternative hypothesis is a hedging-motivated short sale.[18] Informed investors short an industry ETF to hedge their long positions on the underlying stocks for which they have private information on firm-specific fundamentals. This *"long-the-stock/short-the-ETF"* strategy enables informed investors to hedge industry risk, thereby, creating isolated positions to load only on the firm-specific information. Such a strategy seems feasible as ETFs are much easier to short than are individual stocks (see Table A.1 in the Appendix for empirical evidence on shorting difficulty).

These hypotheses have different predictions about the impact of sizable short interest in industry ETFs. The speculation hypothesis predicts a bearish outlook of the industry ETF, so underlying stocks of the industry ETF are likely to underperform. The hedging hypothesis yields the opposite prediction: The large short position on an industry ETF reflects informed investors with optimistic firm-specific information hedging to isolate their positions from the industry risk. Underlying stocks of the industry ETF are likely to outperform.

To disentangle the hedging and speculation hypothesis, we construct a quarterly measure that captures the earnings performance of each ETF's constituent stocks. The measure is the ratio of stocks reporting positive earnings in an ETF, namely the positive earnings ratio (PosSUE). To construct the measure, we define a stock to have positive earnings if its SUE

[18] Note that we do not view these hypotheses as mutually exclusive. Some traders may use industry ETFs to speculate, others to hedge.

is in the top 25% of the entire earnings announcement sample. At every quarter, we then compute the ratio of underlying stocks in an ETF that have positive earnings. The positive earnings ratio captures the likelihood that an ETF's underlying stock outperforms in the future. Panel D of Table 1 reports the summary statistics of the PosSUE for industry and non-industry ETFs.

We use the following regression to test predictions on the speculation and hedging hypothesis:

$$\begin{aligned} PosSUE_{i,t} = {} & \alpha_i + \alpha_t + \beta_1 SIR_{i,t-1} + \beta_2 SIR_{i,t-1} \times DummyCrisis_t \\ & + \beta_3 DummyCrisis_t + controls + \epsilon_{i,t}, \end{aligned} \tag{1}$$

$$\begin{aligned} PosSUE_{i,t} = {} & \alpha_i + \alpha_t + \theta_1 SIR_{i,t-1} + \theta_2 SIR_{i,t-1} \times DummyIndetf_i \\ & + \theta_3 SIR_{i,t-1} \times DummyCrisis_t + \theta_4 SIR_{i,t-1} \times DummyIndetf_i \\ & \times DummyCrisis_t + \theta_5 DummyCrisis_t + \theta_6 DummyCrisis_t \\ & \times DummyIndetf_i + controls + \epsilon_{i,t}. \end{aligned} \tag{2}$$

$PosSUE_{i,t}$ is the positive earnings ratio for ETF i at quarter t and $SIR_{i,t-1}$ is the lagged short interest ratio for i. $DummyCrisis_t$ is a dummy variable for the crisis period, which equals 1 for the period between the fourth quarter of 2006 and the fourth quarter of 2008. The interaction between $DummyCrisis_t$ and $SIR_{i,t-1}$ captures the different predicting power of the short interest ratio in different states of the economy. $DummyIndetf_i$ is a dummy variable which equals 1 if ETF i is an industry ETF. We estimate equation (1) on industry and non-industry ETFs, respectively, and estimate equation (2) on all ETFs. In equation (2), the dummy variable $DummyIndetf_i$ interacting with the $SIR_{i,t-1}$ compares the predicting power of the SIR between industry and non-industry ETFs. In both regressions, we include the log total net asset value of ETF i in the contemporaneous quarter as a control variable. We also control for the year and quarter fixed effect. Due to the high persistence in the SIR for industry ETFs (see Table 2), we further control for the ETF fixed effect. Standard errors are clustered by year-quarter. We estimate equation (1) and (2) on all earnings quarters ranging from 1999 to 2017 and show the regression result in Table 3.

[Insert Table 3 Here]

Table 3 shows that conditioning on the state of the economy reveals different uses of industry ETFs. During the non-crisis period ($DummyCrisis = 0$), for the industry ETF, we find the statistically significant result that lagged SIR positively predicts PosSUE, consistent with the hedging hypothesis. We find the opposite (or insignificant) result for the non-industry ETF. Large short interest predicting more positive earnings is consistent with the *"long-the-stock/short-the-ETF"* strategy carried out by informed investors. The positive predictability of the SIR among industry ETFs suggests this hedging hypothesis is the dominant explanation of sizable short interest during the non-crisis period.

In the crisis period, we find that short interest on industry ETFs reflects more speculative purposes. The coefficient of the interaction term $SIR_{i,t-1} \times DummyCrisis_t$ is significantly negative (-0.074) when we estimate equation (1) on industry ETFs. Our finding is consistent with Karmaziene and Sokolovski (2015) who find during the 2007-2008 crisis that short interest on financial sector ETFs spiked when the financial sector faltered.

Further analysis (reported in Table A.3 in the Appendix) shows that our findings in Table 3 are robust to using the earnings surprises based on the analyst forecast, the industry ETF sample without the requirement of 30 constituents, and different cutoffs for defining the positive/negative SUE. Meanwhile, Panel D of Table A.3 shows that the short interest on industry ETFs can not predict negative earnings surprises, during the non-crisis period, but can positively predict negative earnings surprises during the crisis period.[19] Since the crisis period is less of a norm, for the rest of our analysis we focus on the non-crisis period excluding 2007 and 2008.

2.2 Predictable returns and short interest in industry ETFs

An implication of hedging-motivated short sales is that short interest would create a temporary price pressure, which will revert in the future. This implies that the SIR of the industry ETF, especially the change of the SIR, should positively predict the industry ETF's future return.

To tests this hypothesis, we use both the Fama and MacBeth (1973) approach and portfolio sorting to test the cross-sectional relation between short interest and returns in industry ETFs. For each month, we regress each industry ETF's return against the change of its SIR to estimate the cross-sectional return predictions of industry ETFs' SIR. We then calculate the time series average of these regression coefficients and test for significance based on the time series standard error adjusted by Newey-West. We also include an augmented regression controlling for the characteristics of the underlying stock of the industry ETF. We start our sample from 2005 for both Fama-Macbeth regressions and portfolio sorting due to the scarcity of industry ETFs in earlier periods.[20] Also, we exclude 2007 and 2008 to filter out the unusual period because of the financial crisis. Results are reported in Table 4.

[Insert Table 4 Here]

[19] The negative earnings surprise is 1 if itis in the bottom 25% of the entire earnings announcement sample and is zero otherwise.
[20] In the Fama-Macbeth regressions, the results are robust to including the early sample.

We find that the change in the short interest ratio (∆SIR) positively predicts the future return for the industry ETF. In contrast, we find the ∆SIR at the member stock level negatively predicts the member stock's future return. This latter stock-level result is consistent with past studies (e.g., Diether, Lee, and Werner, 2009; Rapach, Riggenberg, and Zhou, 2016). Our Fama-MacBeth regression result is thus consistent with the hedging hypothesis. The *"long-the-stock/short-the-ETF"* strategy from informed investors creates extreme short selling pressure resulting in a temporary price impact.

In the above test, we use the change in the short interest ratio (∆SIR) rather than the level of the short interest ratio (SIR). The reason is the high persistence of the SIR among industry ETFs (see Table 2). If we sort industry ETFs into five quintiles based on the SIR every month, we find that an ETF in the top quintile in the current month has about 90% chance to remain in the top quintile next month. If we sort on the ∆SIR, the likelihood of remaining in the top portfolio drops down to less than 25%. To avoid the ETF fixed effect coming from the high persistence in the SIR, we use the ∆SIR in the return predictability test, which is also commonly used in the literature (e.g., Yawen, Massa and Zhang, 2016; Chen, Da and Huang, 2018).

We also construct a long-short portfolio to test the positive return predictability of short interest on industry ETFs. At each month, we sort industry ETFs into three groups based on their ∆SIR.[21] Then, we use an equal-weighted scheme to construct the portfolio which longs the ETF in the highest ∆SIR group and shorts the ETF in the lowest ∆SIR group. We hold this portfolio for one month and do monthly rebalancing. The performance of this long-short portfolio is reported in Table 5.

[21] We choose to sort industry ETFs into three groups rather than 10 deciles because of the limited number of industry ETFs in our sample. We have 121 industry ETFs in total until recent years and the early sample has only a few industry ETFs. Thus, portfolio sorting based on deciles make the number of ETFs in each decile to be small.

[Insert Table 5 Here]

Our long-short portfolio generates a four-factor alpha of around 23 basis points per month, and it is statistically significant. We apply a similar approach on stocks, which are members of industry ETFs. In contrast to the ETF-level result, when the test assets are member stocks and the sorting variable is the stock-level ΔSIR, the long-short portfolio with those stocks generates a monthly four-factor alpha of negative 50 basis points.

As a robustness check, we also test the prediction of the ΔSIR on the change of the net asset value (NAV) on industry ETFs. We find consistent result with the return predictability, i.e., ΔSIR of an industry ETF positively predicts its NAV change (see Table A.4). We also find that these return patterns do not revert in the future (see Panel C of Table A.4).

These return patterns, taken at face value, are consistent with the hypothesis that informed investors use the shorts in industry ETFs to hedge, but there is an alternative way to think about the evidence. Specifically, it is possible that the high short interest in industry ETFs are used by arbitragers in correcting mispricing between ETFs and the member stocks, rather than used in the *"long-the-stock/short-the-ETF"* strategy. Although there is no clear theoretical foundation why the shorts in ETF arbitraging activities can positively predict returns, we carry out some tests to rule out this possibility. We show in Panel A of Table A.5 that the ΔSIR does not predict the industry ETF's mispricing (price discount/premium against NAV), which is a common measure of arbitrage activity (e.g., Evans et al 2018). We also control the percentage change of ETF shares outstanding, which is a proxy for ETF arbitrage activity (Brown, Davies and Ringgenberg, 2018).

The positive return predictability is consistent with the hypothesis of hedging-motivated short sales. Extreme short interest reflects informed investors' hedging needs creating a temporary shock, which leads to future price reversion. In contrast to hedging-motivated short sales, speculation-motivated sales should negatively predict the return of the corresponding industry ETF, as it reflects the bearish expectation on the industry. As shown by Table 3, the speculation-motivated short sale dominates in the crisis period. Hence, we redo our long-short portfolio analysis for the crisis period. We find that ΔSIR long-short portfolio returns in the crisis period are significantly lower than in the non-crisis period, with the sign flipped (see Panel B and C in Table A.6). Clearly, the main driver behind short interest on industry ETFs depends on the state of the economy. We do not observe such a pattern at the stock-level.[22]

3. Long the stock and short the ETF

So far, we have shown short interest on industry ETFs reflects the hedging leg of informed investors' *"long-the-stock/short-the-ETF"* strategy in the non-crisis period. We now turn to investigating the channel through which this trading occurs. As noted earlier in the industry report, hedge funds are active users on shorting ETFs and so seem likely candidates to implement this hedging strategy. To explore such a possibility, we examine the contemporaneous correlation between hedge funds trading and industry ETFs' short interest.

3.1 Hedge funds trading and industry ETFs short interest

Based on the data on aggregate hedge funds holdings, we run the following regression model on underlying stocks of industry ETFs

$$AHF_{i,s,t} = \alpha_i + \alpha_t + \beta_1^H DummyPosSUE_{i,s,t+1} + \beta_2^H SIR_{i,t} + \beta_3^H SIR_{i,t} \times DummyPosSUE_{i,s,t+1} + controls + \epsilon_{s,t}. \quad (3)$$

[22] Short interest on the underlying stock of the industry ETF predicts negative returns in both the non-crisis and crisis period (see Table A.7).

For an industry ETF i, $AHF_{i,s,t}$ is hedge funds' abnormal holdings on i's underlying stock s at quarter t. $DummyPosSUE_{i,s,t+1}$ is a dummy variable which takes 1 if stock s reports a positive SUE (i.e., ranks top 25% in our earnings sample) at the following quarter, $t+1$. The interaction term $SIR_{i,t} \times DummyPosSUE_{i,s,t+1}$ captures the conditional correlation between ETF-level short interest and stock-level hedge funds abnormal holdings conditioning on if the stock has positive firm-specific information (positive SUE in the coming announcement). For control variables, we include stock-level market capitalization, book-to-market ratio, institutional ownership, past returns, earnings volatility, and earnings persistence at quarter t. In addition, we control for the year, quarter, and ETF fixed effect. All standard errors are clustered by ETF and year-quarter. Equation (3) is estimated for all underlying stocks of industry ETFs over a sample period covering earnings quarters from 1999 to 2017. We report the regression result in Table 6.

[Insert Table 6 Here]

We find that the coefficient of the interaction term is positively significant. This indicates that short interest on industry ETFs increases with hedge funds abnormal holdings on constituents with positive firm-specific information (positive earnings surprise). On other constituent stocks (e.g., those with negative earnings surprise), short interest on ETFs does not appear to correlate with hedge funds abnormal holdings. Our result suggests that hedge funds trading has a conditional contemporaneous correlation with short interest on industry ETFs. The correlation is conditional on stocks with positive firm-specific fundamentals.

The conditional contemporaneous correlation between hedge funds trading and ETF short interest provides direct evidence on *"long-the-stock/short-the-ETF"* strategy. Furthermore, our finding that the correlation exists conditionally on stocks with positive firm-specific information suggests that short interest is not reflecting ETF mispricing-arbitrage. In mispricing- arbitrage, hedge funds short an overpriced ETF and unconditionally long all its constituents to create the ETF unit for covering up the short position. We do not find that hedge funds increase holdings universally on all underlying stocks when short interest on ETFs increases. Also, we have discussed in the previous section that short interest on industry ETFs does not predict ETF mispricing such as the price discount/premium against the NAV (see Panel A of Table A.5). Moreover, consistent with the finding that short sales are more speculative during the crisis period in Table 3, we find that the industry ETFs' short interest behaves differently from that during the non-crisis period (see Table A.6). In contrasts, the stocks' short interest behaves similar between the crisis and non-crisis periods (see Table A.7).

Could this hedging behaviour also be the norm for other institutional investors? To address this possibility, we use mutual funds, banks, and other institutional investors abnormal holdings as a placebo test for equation (3). We report these result in the rightmost column in Table 6. We do not find any significant correlation between non-hedge funds trading and ETF short interest. The result further confirms hedge funds are most likely to implement the *"long-the-stock/short-the-ETF"* strategy.[23]

3.2 Improve short interest return predictability with hedge funds trading

In Section 2, we showed that hedging-motivated short sales positively predict the industry ETF's return. In the previous subsection, we identified hedge funds as the primary investors who short industry ETFs to hedge. Combining these two findings, we conjecture

[23] In the Appendix, we have another placebo test on the contemporaneous correlation between hedge funds trading and short interest on industry ETFs. We re-estimate equation (3) on the crisis period and report the result in Table A.8. As argued by Chen, Da and Huang (2018), the arbitrage capital is low during the crisis period and thus the long-short strategy is used less frequently. Consistent with the argument by Chen, Da and Huang (2018), we do not find any significant correlation between hedge funds trading and short interest on industry ETFs during that period. That is, we do not find evidence on *"long-the-stock/short-the-ETF"* strategy being implemented during the crisis.

the return predictability of short interest to be stronger if hedge funds holdings on underlying stocks increase contemporarily with the increase of short selling (on the industry ETF).

To test our conjecture, we define a measure called $PosAHF_{i,t}$ for ETF i. $PosAHF_{i,t}$ is a ratio defined as the number of underlying stocks with positive abnormal hedge funds holdings at month t divided by the total number of members in ETF i. $PosAHF_{i,t}$ captures the likelihood of hedge funds increasing their holdings on the underlying stock of ETF i. After that, we add $PosAHF_{i,t}$ and $\Delta SIR_{i,t} \times PosAHF_{i,t}$ to the Fama-MacBeth analysis on the return predictability of ΔSIR of the industry ETF (in Table 4). The interaction term captures the *"long-the-stock/short-the-ETF"* strategy. We report the result in Table 7.

[Insert Table 7 Here]

We find that $\Delta SIR_{i,t} \times PosAHF_{i,t}$ positively predicts the ETF's return. The predictability is statistically significant. As the interaction term better captures short interest related to hedging, Table 7 strengthens our previous conclusion that hedging-motivated short sale positively predicts returns of industry ETFs (see Table 4 and 5). On a related note, we find the predictability of ΔSIR alone switches to negative once the interaction term is added. The result suggests short interest on industry ETFs also involves speculative short selling. This occurs when hedge funds abnormal holdings of industry ETFs' underlying stocks do not increase contemporaneously with short interest on industry ETFs.

The difference between hedging- and speculation-motivated short sale becomes even more apparent when we double sort industry ETFs on ΔSIR (into three groups) and PosAHF (into three groups). In Panel B of Table 7, we see the monthly four-factor alpha of the long-short portfolio based on the ΔSIR rank increases to 50 basis points when we condition on high PosAHF. High PosAHF indicates stronger hedging needs. Conversely, the monthly four-factor alpha becomes negative and insignificant when we condition on low PosAHF. These results further strengthen our argument that informed investors (e.g., hedge funds) short industry ETFs to hedge their long positions in stocks.

3.3 Hedge funds trading performance

The ultimate goal for hedge funds to implement the *"long-the-stock/short-the-ETF"* strategy is to isolate their positions in order to profit from the positive firm-specific information. As a result, jointly considering hedge funds abnormal holdings and short interest on industry ETFs should predict positive stock returns.

To test this hypothesis, we run the following regression on underlying stocks of industry ETFs

$$Ret_{i,s,t+1} = \alpha_i + \alpha_t + \beta_1^S DummyPosAHF_{i,s,t} + \beta_2^S SIR_{i,t} + \beta_3^S DummyPosAHF_{i,s,t} \times SIR_{i,t} + controls + \epsilon_{s,t}, \tag{4}$$

where $Ret_{i,s,t+1}$ is the return of ETF i's member stock s in month $t+1$, $DummyPosAHF_{i,s,t}$ equals to 1 if hedge funds abnormal holdings on stock s is greater than 0% (or 0.5%) in recent quarters prior month t +1. The interaction term $DummyPosAHF_{i,s,t} \times SIR_{i,t}$ captures the joint consideration of hedge funds trading and industry ETFs' short interest. For control variables, we include stock characteristics such as the past 12-month return, market capitalization, book-to-market ratio, asset growth, operating profitability, gross profitability, investment growth, net issuance, accruals, and net operating assets. We also control for the year-quarter, stock, and ETF fixed effect. All standard errors are clustered by ETF and year-quarter.

[Insert Table 8 Here]

In Panel A of Table 8, we find $DummyPosAHF_{i,s,t} \times SIR_{i,t}$ significantly predict positive returns for underlying stocks of industry ETFs. Given that we have controlled for $DummyPosAHF_{i,s,t}$ in the regression, the return predictability from $DummyPosAHF_{i,s,t} \times SIR_{i,t}$ highlights the incremental predicting power of jointly considering the long- and short-side of hedge funds trading.[24] We further examine the channel of the return predictability by asking whether the joint consideration of hedge funds trading and industry ETFs' short interest also predicts the stock fundamental (e.g., earnings surprises or ROA), which can be considered as the sources of return predictions in Table 8. To keep our discussion succinct, we report the details in the Appendix (see Table A.9). The return predictability of the interaction term unveils the information nature of the hedge funds *"long-the-stock/short-the-ETF"* strategy, which is aimed to profit from information on the positive firm-specific fundamentals.

To strengthen our argument, we re-estimate equation (4) on non-hedge funds trading as a placebo test (Panel B of Table 8). We do not find the interaction term between non-hedge funds trading and industry ETFs' short interest predicts stock returns.

4. Implications for market efficiency

Combining hedge funds trading with short interest on industry ETFs, we identified the potential informed trading behind the *"long-the-stock/short-the-ETF"* strategy. As suggested by the theoretical literature in financial innovation, we conjecture that by facilitating informed investors' risk hedging the financial innovation of industry ETFs would improve market efficiency. To test this conjecture, we focus on industry ETFs and earnings announcement events.

[24] Jiao, Massa, and Zhang (2016) also show the joint consideration of hedge funds long and short positions on stocks improves hedge funds' performance. Different from their analysis, we consider short positions on industry ETFs and hedge funds long on stocks.

4.1 *"Long-the-stock/short-the-ETF"* and the PEAD

We examine the impact of hedge fund trading and industry ETFs' short interest on the post-earnings announcement drift (PEAD). If hedging through industry ETFs increases informed trading and thus improve market efficiency, we should expect the *"long-the-stock/short-the-ETF"* strategy reduces PEAD. The reduction of PEAD is direct evidence of market efficiency improvement. As a result, we set out our regression model as follows

$$CAR(1,k)_{i,s,t+1} \tag{4}$$

$$= \alpha_i + \alpha_t + \beta_1^C SUERank_{i,s,t+1} \times SIR_{i,t} \times DummyPosAHF_{i,s,t} + \beta_2^C SUERank_{i,s,t+1} \times SIR_{i,t} + \beta_3^C SUERank_{i,s,t+1} \times DummyPosAHF_{i,s,t} + \beta_4^C SIR_{i,t} \times DummyPosAHF_{i,s,t} + \beta_5^C SUERank_{i,s,t+1} + \beta_6^C SIR_{i,t} + \beta_7^C DummyPosAHF_{i,s,t} + controls + \epsilon_{s,t}.$$

$CAR(1,k)_{i,s,t+1}$ is the cumulative size-adjusted return from the first to the kth post-earnings-announcement trading day in quarter $t+1$ for ETF i's member stock s. $SUERank_{i,s,t+1}$ is the quintile ranking of SUE for all stocks in our earnings announcement sample. The triple interaction term, $SUERank_{i,s,t+1} \times SIR_{i,t} \times DummyPosAHF_{i,s,t}$, captures the joint impact of hedge fund trading and industry ETFs' short interest on the PEAD. For control variables, we include market capitalization, book-to-market ratio, past 12-month returns, and the interaction between them and $SUERank_{i,s,t+1}$. We also control for the year-quarter, stock, and ETF fixed effect. All standard errors are clustered by stock and announcement date. We run the regression on all stocks in our earnings announcement sample ranging from 1999 to 2017 except for the crisis period. Table 9 reports our regression results.

[Insert Table 9 Here]

We find the coefficient on $SUERank_{i,s,t+1} \times SIR_{i,t} \times DummyPosAHF_{i,s,t}$ is significantly negative suggesting a reduction effect on the PEAD. Thus, when short interest on industry ETF i increases and hedge funds abnormally increase their holdings on i's member stock s, there is less drift in the subsequent earnings announcement of stock s. This reduction effect holds regardless of the choice of the post-earnings window.

We also note that the interaction term $SUERank_{i,s,t+1} \times SIR_{i,t}$ is positive, whereas combining the point estimate of it with $SUERank_{i,s,t+1} \times SIR_{i,t} \times DummyPosAHF_{i,s,t}$ is negative. This result suggests hedge funds trading helps to disentangle hedging- and speculation-motivated short interest on industry ETFs. The triple interaction term captures the *"long-the-stock/short-the-ETF"* strategy, clearly showing that hedging-motivated short sales reduce the PEAD.

4.2 *"Short-the-stock/long-the-ETF"* and the negative SUE

Until now, we focused on the *"long-the-stock/short-the-ETF"* strategy to bet on positive firm-specific fundamentals. In theory, with the strategy of *"long-the-ETF/short-the-stock"*, industry ETFs can also be used as a hedge to bet on negative firm-specific fundamentals. In this subsection, we explore the possibility of this reverse strategy by examining the correlation between hedge funding trading on industry ETFs and stock-level short interest, conditioning on the negative SUE. We note at the outset that the higher shorting costs or limit-to-arbitrage costs (e.g., liquidity or idiosyncratic volatilities) associated with individual stocks makes this alternative strategy more expensive, suggesting that it may not be as feasible to implement.

We use a regression model in the spirit of equation (3) to study the contemporaneous correlation between stock-level short interest and ETF-level hedge funds trading. The main difference is we focus on negative firm-specific fundamentals.

$$\Delta SIR(S)_{i,s,t} = \alpha_i + \alpha_t + \beta_1^N DummyNegSUE_{i,s,t+1} + \beta_2^N AHF(E)_{i,t} + \beta_3^N AHF(E)_{i,t} \times DummyNegSUE_{i,s,t+1} + controls + \epsilon_{s,t}. \quad (5)$$

$\Delta SIR(S)_{i,s,t}$ is the change of the short interest ratio for stock s at quarter t. Stock s is a member of industry ETF i. We use the change of the short interest ratio to avoid persistence in the SIR.[25] $DummyNegSUE_{i,s,t+1}$ is a dummy variable indicating negative SUE for stock s in the following quarter $t+1$. In correspondence to our definition of the positive SUE, the negative SUE is defined as the bottom 25% SUE in our earnings announcement sample. The interaction term $AHF(E)_{i,t} \times DummyNegSUE_{i,s,t+1}$ captures the contemporaneous correlation between stock-level short interest ($\Delta SIR(S)_{i,s,t}$) and ETF-level hedge funds trading ($AHF(E)_{i,t}$) conditional on the stock having negative information ($DummyNegSUE_{i,s,t+1}$). We include market capitalization, book-to-market ratio, institutional ownership, past returns, earnings volatility, and earnings persistence as our control variables. We also control for the year, quarter, and ETF fixed effect. All standard errors are clustered by ETF and year-quarter. Table 10 reports our regression results.

[Insert Table 10 Here]

We find the coefficient of the interaction term $AHF(E)_{i,t} \times DummyNegSUE_{i,s,t+1}$ is significantly positive. The positive coefficient suggests when the member stock has negative firm-specific fundamentals, we find hedge funds increase their holdings on the industry ETF contemporarily with short interest of the stock increases. Further, we do not find evidence on the unconditional correlation between ETF-level hedge funds trading and stock-level short interest. The existence of the conditional contemporaneous correlation is consistent with the

[25] An alternative approach is to use SIR and control for the stock fixed effect.

"long-the-ETF/short-the-stock" strategy being used to profit from negative firm-specific fundamentals.

When comparing the result in Table 10 with Table 6, we find the conditional contemporaneous correlation between hedge funds trading and short selling is more apparent (larger *t-stat.*) with positive firm-specific information. This asymmetry could be due to the higher cost of shorting individual stocks, noted earlier. To investigate this possibility, we split our stocks into large and small stocks, then redo the estimation of equation (5). Since large stocks tend to have lower shorting costs, we conjecture the conditional contemporaneous correlation exists primarily among large stocks.[26] Panel A of Table A.10 in the Appendix confirms our conjecture. Panel B of Table A.10 further shows the *"long-the-ETF/short-the-stock"* strategy predicts negative stock returns mostly among large stocks.

4.3 Industry ETFs and market efficiency

Up to this point, we have shown industry ETFs are used to hedge industry risk. Informed investors like hedge funds can either apply the *"long-the-stock/short-the-ETF"* strategy for positive firm-specific information or the *"long-the-ETF/short-the-stock"* strategy for negative firm-specific information. With informed investors better hedged, they trade more aggressively leading to more information being impounded into the market, and so greater market efficiency. In this last section, we directly test if industry ETFs improve market efficiency. Again, we first focus on PEAD to test the market efficiency.

Our regression model is the following

[26] An alternative approach to examine the constraint of shorting cost on taking the *"long-the-ETF/short-the-stock"* strategy is to sort stocks based on the difficulty to short measure in the Markit data. However, this approach is not ideal due to poor coverage of Markit in the earlier period.

$$CAR(1,k)_{s,t+1} \tag{6}$$
$$= \alpha_s + \alpha_t + \beta_1^S SUERank_{s,t+1} + \beta_2^S DummyMember_{s,t}$$
$$+ \beta_3^S SUERank_{s,t+1} \times DummyMember_{s,t} + controls + \epsilon_{s,t}.$$

$CAR(1,k)_{s,t+1}$ and $SUERank_{s,t+1}$ are defined similarly as in equation (4). $DummyMember_{s,t}$ is dummy variable indicating if stock s belongs to any industry ETF. The interaction term $SUERank_{s,t+1} \times DummyMember_{s,t}$ captures the industry membership effect on the PEAD. We control for market capitalization, book-to-market ratio, and interact them with $SUERank_{s,t+1}$. Also, we control for the year-quarter and industry fixed effect. All standard errors are clustered by stock and announcement date.

[Insert Table 11 Here]

We are aware that stocks are not randomly included in ETFs. There could be factors driving the inclusion process that mixes up with the industry ETF membership effect on the PEAD. To control for the confounding effect, we apply a propensity score matching procedure based on the industry category, firm size, and book-to-market ratio. Panel A and B of Table 11 show the outcome of our propensity score matching. In Panel A, we indeed find that ETF member stocks have different firm characteristics from non-member stocks. For example, ETF member stocks tend to be larger. After the propensity score matching (Panel B of Table 11), there are no significant differences in firm characteristics between ETF member and non-member stocks.

Panel C of Table 11 shows our regression results on equation (6) on the matched sample over the same period as in equation (4). We find the interaction term $SUERank_{s,t+1} \times DummyMember_{s,t}$ is negatively correlated with $CAR(1,k)_{s,t+1}$. The negative correlation suggests the industry ETF membership significantly reduces the PEAD of the stock. The reduction effect indicates industry ETFs improve market efficiency.

In the Appendix, we include the regression result of equation (6) on all stocks in our earnings announcement sample, i.e., the full sample without the propensity score matching. We find consistent but stronger results (see Table A.11) for the full sample before the propensity score matching. This suggests the propensity score matching is needed so the effect is not inflated. Also, we conduct a placebo test by replacing the industry ETF membership with the non-industry ETF membership in equation (6). We do not expect that non-industry ETFs are used for hedging, since short interest on non-industry ETFs is quite small. Thus, we do not expect to find a similar impact from non-industry ETFs. Table A.12 confirms our conjecture, as we do not find the non-industry ETF membership reduces the PEAD.

In addition to PEAD, we also consider alternative measures of price efficiency, including the price delay measure and the variance ratio, based on the matched samples. The results are reported in Table A.13. In short, results on alternative tests are consistent with the result on the PEAD test.

5. Conclusion

Can industry ETFs facilitate informed trading and enhance the informational efficiency of the market? Our results show that they can by facilitating the hedging of industry risk for informed investors. We find sizeable short interest on industry ETFs during the non-crisis period is associated with the *"long-the-stock/short-the-ETF"* strategy implemented by informed investors, such as hedge funds. We find that the change in industry short interest predicts the future return for the industry ETF. This return predictability is consistent with the hedging-based use of the ETF inducing a temporary price impact. Using earnings announcements, we show that the *"long-the-stock/short-the-ETF"* strategy reduces the post-earnings-announcement-drift in response to the earnings surprise. We also show, on a propensity score match sample, the same effect holds for the industry ETF membership. Overall, industry ETFs appear to be a valuable innovation in the market.

One aspect of our results that we find particularly intriguing is the asymmetry of the effects: the effect for the *"long-the-stock/short-the-ETF"* strategy is stronger than the reverse strategy, the *"long-the-ETF/short-the-stock"* strategy, aiming at negative firm-specific information. We believe this reflects another aspect of this financial innovation as industry ETFs reduce the transactions cost of shorting, making the "*long-the-stock/short-the-ETF*" strategy feasible. No similar innovation exists to reduce the shorting costs of individual stocks, but perhaps future financial innovation can address this problem.

References

Allen, Franklin, and Douglas Gale, 1994, *Financial Innovation and Risk Sharing* (MIT Press).

Asquith, Paul, Parag A. Pathak, and Jay R. Ritter, 2005, Short interest, institutional ownership, and stock returns, *Journal of Financial Economics* 78, 243–276.

Battalio, Robert, and Paul Schultz, 2011, Regulatory uncertainty and market liquidity: The 2008 short sale ban's impact on equity option markets, *Journal of Finance* 66, 2013–2053.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, 2011, Hedge fund stock trading in the financial crisis of 2007–2009, *Review of Financial Studies* 25, 1–54.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, 2018, Do ETFs increase volatility? *Journal of Finance* 73, 2471–2535.

Bhattacharya, Ayan, and Maureen O'Hara, 2017, Can ETFs increase market fragility? Effect of information linkages in ETF markets, *Working Paper*, 1–52.

Boehmer, Ekkehart, Charles M. Jones, and Xiaoyan Zhang, 2008, Which shorts are informed? *Journal of Finance* 63, 491–527.

Brown, David, Shaun William Davies, and Matthew C. Ringgenberg, 2018, ETF Arbitrage and Return Predictability, *Working Paper*.

Brunnermeier, Markus K., and Stefan Nagel, 2005, Hedge funds and the technology bubble, *Journal of Finance* 59, 2013–2040.

Chen, Yong, Zhi Da, and Dayong Huang, 2018, Arbitrage trading: The long and the short of it, *Review of Financial Studies* 66, 341.

Chen, Zhiwu, 1995, Financial innovation and arbitrage pricing in frictional economies, *Journal of Economic Theory* 65, 117–135.

Cong, Lin William, and Douglas Xun Xu, 2016, Rise of factor investing: Asset prices, informational efficiency, and security design, *Working Paper*.

Da, Zhi, and Sophie Shive, 2017, Exchange traded funds and asset return correlations, *European Financial Management* 24, 136–168.

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

Diamond, Douglas W., and Robert E. Verrecchia, 1987, Constraint on short-selling and asset price adjustment to private information, *Journal of Financial Economics* 18, 277–311.

Diether, Karl B., Kuan-Hui Lee, and Ingrid M. Werner, 2009, Short-sale strategies and return predictability, *Review of Financial Studies* 22, 575–607.

Dow, James, 1998, Arbitrage, hedging, and financial innovation, *Review of Financial Studies* 11, 739–755.

Duffie, Darrell, and Rohit Rahi, 1995, Financial market innovation and security design: An introduction, *Journal of Economic Theory* 65, 1–42.

Evans, Richard B., Rabih Moussawi, Michaels S. Pagano, and John Sedunov, 2018, ETF Short Interest and Failures-to-Deliver: Naked Short-Selling or Operational Shorting?, *Working Paper.*

Fama, Eugene F., and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Easley, David, David Michayluk, Maureen O'Hara, and Talis J. Putnins, 2018, The active world of passive investing, *Working Paper*.

Glosten, Lawrence, Suresh Nallareddy, and Yuan Zou, 2017, ETF activity and informational efficiency of underlying securities, *Working Paper*.

Griffin, John M., and Jin Xu, 2009, How smart are the smart guys? A unique view from hedge fund stock holdings, *Review of Financial Studies* 22, 2531–2570.

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

Israeli, Doron, Charles M. C. Lee, and Suhas A. Sridharan, 2017, Is there a dark side to exchange traded funds? An information perspective, *Review of Accounting Studies* 22, 1048–1083.

Jiang, Wenxi, 2017, Leveraged speculators and asset prices, *Working Paper*, 1–59.

Jiao, Yawen, Massimo Massa, and Hong Zhang, 2016, Short selling meets hedge fund 13F: An anatomy of informed demand, *Journal of Financial Economics* 122, 544–567.

Li, Frank Weikai, and Qifei Zhu, 2017, Short selling ETFs, *Working Paper*, 1–68.

Livnat, Joshua, and Richard R. Mendenhall, 2006, Comparing the post–earnings announcement drift for surprises calculated from analyst and time series forecasts, *Journal of Accounting Research* 44, 177–205.

Madhavan, Ananth, and Aleksander Sobczyk, 2015, Price dynamics and liquidity of exchange-traded funds, *Working Paper*.

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

Rapach, David E., Matthew C. Ringgenberg, and Guofu Zhou, 2016, Short interest and aggregate stock returns, *Journal of Financial Economics* 121, 46–65.

Simsek, Alp, 2013a, Financial innovation and portfolio risks, *American Economic Review: Papers and Proceedings* 103, 398–401.

Simsek, Alp, 2013b, Speculation and risk sharing with new financial assets, *Quarterly Journal of Economics* 128, 1365–1396.

Wermers, Russ, and Jinming Xue, 2015, Intraday ETF trading and the volatility of the underlying, *Working Paper*.

Table 1: Summary statistics

Panel A and B report the summary statistics on quarter short interest ratio (*SIR*), log of shares outstanding, price, volume, and total net asset value (*TNA*) for industry and non-industry ETFs, respectively. The quarter measure is constructed by taking the average of monthly observations. Panel C reports the summary statistics for stocks in our earnings announcement sample excluding the crisis period from the fourth quarter of 2006 to the fourth quarter of 2008. *DummyIndetfown* is the dummy variable which equals to 1 if the stock is a member of an industry ETF. *CAR* is the -1 to +1 cumulative abnormal daily return around the earnings announcement date based on the Fama-French-Carhart four factor model. *SUE* is the standardized earnings surprise computed from a rolling seasonal random walk model. Both holdings are standardized by shares outstanding. *log(MktCap)* is the log transformed market capitalization. *BM* is the book-to-market ratio where the book value is measured as the preceding fiscal year, and market value is measured as of the end of that calendar year. *TR* is the turnover measured as the average of the daily ratios of volume over shares outstanding from -40 to -11 of each announcement. *MOM* is the cumulative raw return over the six-month period ending one month before the announcement month. *EarnPerst* is the earnings persistence as of the first-order auto-regressive coefficient of quarterly earnings over the past four years. *NumEst* is the number of analysts. Panel D reports the summary statistics on the ratio of positive *SUE* in an ETF over our sample period. The positive *SUE* is defined as the *SUE* exceeding the 75 percentile of all *SUE*s in the sample.

Panel A: Industry ETFs

	Mean	**Std.**	**5%**	**25%**	**50%**	**75%**	**95%**
SIR	**0.115**	0.209	0.001	0.007	0.022	0.112	**0.594**
Log(Shrout)	15.694	1.682	12.812	14.670	15.703	16.781	18.455
Price	53.919	34.132	16.875	29.637	47.242	67.529	118.782
Log(Dollar Volume)	13.879	2.347	10.255	12.297	13.756	15.116	18.273
Log(TNA in $ millions)	5.660	1.769	2.588	4.516	5.771	6.849	8.621

Panel B: Non-industry ETFs

	Mean	**Std.**	**5%**	**25%**	**50%**	**75%**	**95%**
SIR	**0.037**	0.111	0.000	0.003	0.007	0.019	**0.170**
Log(Shrout)	15.487	1.834	12.612	14.197	15.278	16.797	18.689
Price	58.196	37.599	16.315	29.497	49.793	76.727	129.401
Log(Dollar Volume)	12.992	2.342	9.715	11.418	12.663	14.337	17.103
Log(TNA in $ millions)	5.511	2.105	2.351	3.978	5.249	6.864	9.349

Panel C: The earnings announcement sample

	Mean	**Std.**	**5%**	**25%**	**50%**	**75%**	**95%**
DummyIndetfown	0.359	0.480	0.000	0.000	0.000	1.000	1.000
CAR	0.001	0.093	-0.137	-0.037	0.000	0.039	0.139
SUE	-0.001	0.060	-0.071	-0.006	0.001	0.007	0.064
log(MktCap)	19.790	2.002	16.755	18.291	19.654	21.121	23.321
BM	0.686	0.589	0.107	0.303	0.538	0.874	1.783
TR	0.006	0.006	0.000	0.002	0.004	0.007	0.018
MOM	0.084	0.378	-0.467	-0.133	0.047	0.236	0.775
EarnPerst	0.260	0.361	-0.265	-0.013	0.213	0.521	0.898
NumEst	3.983	5.452	0.000	0.000	2.000	5.000	16.000

Panel D: Panel D: The ratio of positive SUE in an ETF

	Mean	**Std.**	**5%**	**25%**	**50%**	**75%**	**95%**
Ind. ETFs	0.164	0.115	0.037	0.088	0.136	0.202	0.400
Non-ind. ETFs	0.152	0.084	0.048	0.098	0.136	0.186	0.309

Table 2: The characteristics and persistence of different short interest categories

Table 2 reports the distribution of industry ETFs/non-industry ETFs and the short interest ratio persistence for each short interest ratio category. The sample period is 2005-2017. Panel A reports the percentages of industry/non-industry ETFs for each quintile of short interest ratio. In Panel A, we pool all ETFs together. At the end of each month, we sort all ETFs into quintiles based on the short interest ratio (*SIR*). We calculate the percentages of industry and non-industry ETFs for each month and Panel A reports the means for percentages of industry and non-industry ETFs in our sample. Panel B focuses on industry ETFs and calculate the Markov transition matrix of short interest ratios. At the end of each month, we sort all industry ETFs into quintiles based on the short interest ratio (*SIR*) and calculate the month-to-month transitional likelihood. Panel B reports the average transitional likelihood.

Panel A: Monthly average percentages of industry and non-industry ETFs in each SIR group

SIR Rank	**Industry ETFs**	**Non-Industry ETFs**
Low	16%	84%
2	21%	79%
3	26%	74%
4	36%	64%
High	59%	41%

Panel B: Markov transition matrix for industry ETF sorting on monthly SIR

SIR Rank	**Low**	**2**	**3**	**4**	**High**
Low	0.60	0.27	0.10	0.03	0.00
2	0.25	0.47	0.23	0.04	0.00
3	0.10	0.22	0.51	0.15	0.01
4	0.03	0.04	0.15	0.68	0.10
High	0.00	0.00	0.01	0.10	0.88

Table 3: Regress the positive earnings ratio on short interest at the ETF level

Table 3 reports the result of regressing the positive earnings ratio (*PosSUE*) on the lagged short interest ratio (*SIR*), i.e.,

$$PosSUE_{i,t} = \alpha_i + \alpha_t + \beta_1 SIR_{i,t-1} + \beta_2 SIR_{i,t-1} \times DummyCrisis_t + \beta_3 DummyCrisis_t + controls + \epsilon_{i,t},$$

$$PosSUE_{i,t} = \alpha_i + \alpha_t + \theta_1 SIR_{i,t-1} + \theta_2 SIR_{i,t-1} \times DummyIndetf_i + \theta_3 SIR_{i,t-1} \times DummyCrisis_t + \theta_4 SIR_{i,t-1} \times DummyIndetf_i \times DummyCrisis_t + \theta_5 DummyCrisis_t + \theta_6 DummyCrisis_t \times DummyIndetf_i + controls + \epsilon_{i,t}.$$

We use a dummy variable (*DummyCrisis*) to capture the 2007–2008 financial crisis effect on the level of earnings, and the interaction variable (*SIR* x *DummyCrisis*) to examine the differential predictability of short interest on earnings between the crisis and non-crisis period. *DummyCrisis* equals to one for the period from the fourth quarter of 2006 to the fourth quarter of 2008, and zero otherwise. In our controls, we include log total net asset value, and the year, quarter, and ETF fixed effect. All standard errors are clustered by year-quarter. We report regression results in industry ETF and non-industry ETF samples separately. In addition, we pool the sample of industry ETF and non-industry ETF and generate a dummy variable, *DummyIndetf*, to identify industry ETFs. In the all-ETF sample, we regress *PosSUE* on *SIR*, *DummyIndetf*, *DummyCrisis* and their interaction terms. *** is significant at 1%, ** is significant at 5%, and * is significant at 10%. t-statistic is reported in the parenthesis.

DepVar: $PosSUE_{i,t}$	**Industry ETFs**	**Non-industry ETFs**	**All ETFs**
$SIR_{i,t-1}$	0.050***	-0.020	-0.022
	(2.75)	(-1.42)	(-1.57)
$SIR_{i,t-1} \times DummyIndetf_i$			0.067***
			(3.70)
$SIR_{i,t-1} \times DummyCrisis_t$	-0.074***	-0.003	-0.003
	(-4.90)	(-0.23)	(-0.18)
$SIR_{i,t-1} \times DummyIndetf_i \times DummyCrisis_t$			-0.071***
			(-4.02)
$DummyCrisis_t$	-0.008	-0.068**	-0.044*
	(-0.35)	(-2.39)	(-1.78)
$DummyCrisis_t \times DummyIndetf_i$			0.006
			(0.99)
Year FE	Yes	Yes	Yes
Qtr FE	Yes	Yes	Yes
ETF FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	4,778	9,638	14,416
Adjusted R-Squared	0.514	0.683	0.597

Table 4: Fama-MacBeth regression of returns on short interest ratios

Table 4 reports the time-series averages of slope coefficients from Fama and MacBeth (1973) cross-sectional regressions on returns and changes in the short interest ratio, *ΔSIR*, for industry ETFs and their member stocks, respectively. t-statistic reported in the parenthesis is calculated using the average slope coefficient divided by its time-series standard error adjusted by Newey-West with one lag. For each industry ETF, we average the member stocks' characteristics and use the average as a control in our regression. In our control variables, we include stock characteristics as of month *t* end, including past 12-month returns, market capitalization, book-to-market ratio, asset growth, operating profitability, gross profitability, investment growth, net issuance, accruals, and net operating assets. We also repeat the Fama-MacBeth regression on industry ETFs' constituent stocks and report in the last two columns. For the stock-level regression, the control variable corresponds to each stock's own characteristics. The sample period of our analysis is reported in the last row of the table. *** is significant at 1%, ** is significant at 5%, and * is significant at 10%.

DepVar	$\boldsymbol{Ret_{t+1}}$ **of industry ETFs**		**DepVar**	$\boldsymbol{Ret_{t+1}}$ **of member stocks**	
ΔSIR_t*(ETF-level)*	0.030***	0.023**	ΔSIR_t*(stock-level)*	-0.106***	-0.104***
	(2.78)	(2.27)		(-3.02)	(-2.87)
Intercept	0.013***	0.015	*Intercept*	0.014***	0.023***
	(3.72)	(1.31)		(4.11)	(3.31)
Controls	No	Yes	*Controls*	No	Yes
Sample Period:	2005.01 - 2006.09, 2009.01 - 2017.12		Sample Period:	1999.01 - 2006.09, 2009.01 - 2017.12	

Table 5: Long-short portfolio sorting on ΔSIR

Table 5 reports the average monthly excess returns, CAPM alpha, Fama and French 3-factor alpha, and Fama-French-Carhart 4-factor alpha for each portfolio and the high-low portfolio based on *ΔSIR*. At the end of each month, all industry ETFs or member stocks are sorted into deciles based on *ΔSIR* in that month. Then, we track the equal-weighted portfolio returns over the next month. Panel A reports results for industry ETFs. Holding periods in Panel A is from January 2005 to December 2017, excluding crisis period: October 2006 to December 2008. Panel B reports results for member stocks. Member stocks with prices below $5 a share or are in the bottom NYSE size decile are excluded from the sample. Holding period in Panel B is from January 1999 to December 2017, excluding crisis period: October 2006 to December 2008. Standard errors are Newey-West adjusted with one lag.

Panel A: Industry ETFs

	Excess returns		CAPM alpha		3-factor alpha		4-factor alpha	
Portfolios	coef.	t-stat	coef.	t-stat	coef.	t-stat	coef.	t-stat
Bottom 30%	1.13	3.07	-0.29	-4.31	-0.23	-3.94	-0.21	-3.84
Mid 40%	1.29	3.72	-0.08	-1.14	-0.04	-0.58	-0.02	-0.41
Top 30%	1.37	3.78	-0.05	-0.54	0.00	0.01	0.01	0.17
Top - Bottom	**0.24**	**2.65**	**0.24**	**2.59**	**0.23**	**2.50**	**0.23**	**2.40**

Panel B: Member stocks of industry ETFs

	Excess returns		CAPM alpha		3-factor alpha		4-factor alpha	
Deciles	coef.	t-stat	coef.	t-stat	coef.	t-stat	coef.	t-stat
1 (low)	1.25	3.15	0.36	1.47	0.01	0.09	0.06	0.38
2	1.28	3.52	0.45	2.12	0.13	0.84	0.18	1.27
3	1.27	3.83	0.52	2.60	0.22	1.53	0.26	1.82
4	1.33	4.11	0.59	3.07	0.28	2.24	0.29	2.39
5	1.20	3.63	0.50	2.29	0.16	1.12	0.18	1.31
6	1.31	4.13	0.61	3.05	0.28	2.38	0.31	2.67
7	1.34	3.95	0.59	2.81	0.25	1.97	0.29	2.21
8	1.11	3.16	0.33	1.51	0.02	0.13	0.04	0.30
9	0.92	2.50	0.10	0.48	-0.22	-1.40	-0.17	-1.14
10 (high)	0.85	1.97	-0.12	-0.50	-0.46	-2.46	-0.39	-2.16
10-1	**-0.40**	**-3.22**	**-0.47**	**-3.97**	**-0.47**	**-3.93**	**-0.45**	**-3.67**

Table 6: Hedge funds abnormal holdings and short interest on industry ETFs

Table 6 reports the regression of hedge funds abnormal holdings (*AHF*) on short interest on industry ETFs (SIR),

$$AHF_{i,s,t} = \alpha_i + \alpha_t + \beta_1^H DummyPosSUE_{i,s,t+1} + \beta_2^H SIR_{i,t} + \beta_3^H SIR_{i,t} \times DummyPosSUE_{i,s,t+1} + controls + \epsilon_{s,t}.$$

AHF is the abnormal holdings by hedge funds of ETF's constituent stocks standardized by stocks' total shares outstanding at the quarter end. Abnormal holdings are estimated as the difference between the current quarter holdings and the moving average of holdings in past four quarters. *SIR* is the quarter end short interest ratio of the ETF. *DummyPosSUE* is a dummy that takes one if the stock's SUE of that quarter ranks top 25% in our sample. *SIR* x *DummyPosSUE* is the interaction term. In our control variables, we include market capitalization, book-to-market ratio, institutional ownership, past returns, earnings volatility, and earnings persistence as of quarter *t* end. In addition, we also control for the year, quarter, and ETF fixed effect. All standard errors are clustered by ETF and year-quarter. We run the above regression model on our quarterly earnings announcements ranging from 1999 to 2017 except for the crisis period (from the fourth quarter of 2006 to the fourth quarter of 2008). As a placebo test, we replace the dependent variable by abnormal holdings of non-hedge funds and re-run the above regression. We report the placebo test result in the rightmost column. *** is significant at 1%, ** is significant at 5%, and * is significant at 10%. t-statistic is reported in the parenthesis.

	HF Abnormal Holding			**Non-HF Abnormal Holding**
$Pos\_SUE_{i,s,t+1}$	0.053***	0.042**	0.039**	0.149**
	(3.25)	(2.63)	(2.50)	(2.37)
$SIR_{i,t}$		0.004	-0.004	0.147
		(0.07)	(-0.07)	(0.63)
$SIR_{i,t} \times Pos\_SUE_{i,s,t+1}$		0.125***	0.145***	0.034
		(3.58)	(3.33)	(0.23)
Log(Mktcap)			-0.035***	-0.138***
			(-4.31)	(-3.53)
BM			0.008	-0.258***
			(0.35)	(-2.81)
Institutional Ownership			0.329***	1.237***
			(3.37)	(5.67)
Reversal			0.050	0.387
			(0.52)	(1.15)
Momentum			-0.041	1.535***
			(-0.80)	(6.86)
Earnings Volatility			0.001**	0.010*
			(2.30)	(1.99)
Earnings Persistence			0.014	0.059
			(0.99)	(1.29)
Year FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
ETF FE	Yes	Yes	Yes	Yes
Observations	372,209	372,209	355,726	354,317
Adjusted R-Squared	0.0141	0.0141	0.0175	0.0644

Table 7: Strategy based on ETFs short interest and hedge funds abnormal holdings

Table 7 reports the improved trading strategy for industry ETFs based on the change of ETF's short interest ratio and hedge funds abnormal holdings of the ETF's underlying stocks. Panel A reports the time-series averages of slope coefficients from Fama and MacBeth (1973) cross-sectional regressions of industry ETF returns in the next month, *Ret*, on the current month's change in *SIR*, the *ΔSIR*, and *PosAHF*, and their interaction term. As of the current month, *PosAHF* is defined as number of holding stocks whose abnormal hedge fund holding is positive in the latest quarter divided by total number of holding stocks at the latest. In our control variables, we include stock characteristics as of the current month, including past 12-month returns, market capitalization, book-to-market ratio, asset growth, operating profitability, gross profitability, investment growth, net issuance, accruals, and net operating assets. t-statistic reported in the parenthesis is calculated using the average slope coefficient divided by its time-series standard error adjusted by Newey-West with one lag. Panel B reports holding period returns of industry ETF portfolios sorted on *ΔSIR* and *PosAHF*. At each month end, we sort industry ETFs into three groups (lowest 30% / middle 40% / highest 30%) based on *ΔSIR* in the month. Within each *ΔSIR* portfolio, we further sort industry ETFs into three groups (low, middle, and high) based on *PosAHF* at the latest quarter end. We hold the portfolios in the next month and report equally-weighted portfolio excess returns, CAPM alpha, Fama-French 3-factor alpha, and Fama-French-Carhart 4-factor alpha. The sample period of our analysis is from January 2005 to December 2017, excluding crisis period (from the fourth quarter of 2006 to the fourth quarter of 2008). *** is significant at 1%, ** is significant at 5%, and * is significant at 10%.

Panel A: Fama-MacBeth regression

DepVar: Ret_{t+1} of industry ETFs				
ΔSIR_t	0.030***		-0.091*	-0.07
	(2.78)		(-1.80)	(-1.49)
$PosAHF_t$		-0.002	0.000	0.003
		(-0.21)	(-0.05)	(0.44)
$\Delta SIR_t \times PosAHF_t$			0.456**	0.311**
			(2.44)	(2.08)
Intercept	0.013***	0.015***	0.014***	0.017
	(3.72)	(4.94)	(4.65)	(1.38)
Controls	Yes	Yes	No	Yes

Panel B: Returns on portfolios sorting on *ΔSIR* and *PosAHF*

	Excess returns			
	High *PosAHF*		Low *PosAHF*	
Portfolio	coef.	t-stat	coef.	t-stat
Low *ΔSIR*	0.84	1.86	1.32	4.05
High *ΔSIR*	1.35	3.13	1.32	4.02
High *ΔSIR* - Low *ΔSIR*	**0.51**	**2.75**	**0.00**	**-0.02**
	CAPM alpha			
	High *PosAHF*		Low *PosAHF*	
Portfolio	coef.	t-stat	coef.	t-stat
Low *ΔSIR*	-0.74	-3.87	0.10	0.80
High *ΔSIR*	-0.20	-0.91	0.05	0.40
High *ΔSIR* - Low *ΔSIR*	**0.54**	**2.58**	**-0.05**	**-0.37**
	3-factor alpha			
	High *PosAHF*		Low *PosAHF*	
Portfolio	coef.	t-stat	coef.	t-stat
Low *ΔSIR*	-0.67	-3.89	0.15	1.21
High *ΔSIR*	-0.13	-0.62	0.09	0.78
High *ΔSIR* - Low *ΔSIR*	**0.54**	**2.52**	**-0.06**	**-0.45**
	4-factor alpha			
	High *PosAHF*		Low *PosAHF*	
Portfolio	coef.	t-stat	coef.	t-stat
Low *ΔSIR*	-0.62	-3.53	0.15	1.17
High *ΔSIR*	-0.10	-0.50	0.09	0.77
High *ΔSIR* - Low *ΔSIR*	**0.52**	**2.34**	**-0.06**	**-0.44**

Table 8: Trading performance of hedge funds and non-hedge funds

Table 8 reports results on using stock-level hedge funds abnormal holdings and ETF-level short interest to predict stock returns. The regression model is the following

$$Ret_{i,s,t+1} = \alpha_i + \alpha_t + \beta_1^S DummyPosAHF_{i,s,t} + \beta_2^S SIR_{i,t} + \beta_3^S DummyPosAHF_{i,s,t} \times SIR_{i,t} + controls + \epsilon_{s,t},$$

Ret is the stock return (the excess return or the Daniel, Grinblatt, Titman, and Wermers, DGTW, characteristics-adjusted return) in the next month. *DummyPosAHF* is a dummy variable, which equals one if abnormal hedge fund holding on stock s is greater than 0% (or 0.5%) in the latest quarter as of the current month. *SIR* is the short interest ratio of ETF at the end of the current month. In our control variables, we include the current month stock characteristics, including past 12-month returns, market capitalization, book-to-market ratio, asset growth, operating profitability, gross profitability, investment growth, net issuance, accruals, and net operating assets. In addition, we also control for the year-quarter, stock, and ETF fixed effect. All standard errors are clustered by ETF and year-quarter. In Panel B, as a placebo test, we replace *DummyPosAHF* by the dummy variable of non-hedge funds abnormal holdings, *PosANHF*, which is constructed using non-hedge funds holdings. *PosANHF* equals one if abnormal non-hedge fund holding is greater than 0% (or 0.5%) in the latest quarter as of the current month. We run the above regression model on all underlying stocks of industry ETFs from 1999 to 2017 except for the crisis period (from the fourth quarter of 2006 to the fourth quarter of 2008). *** is significant at 1%, ** is significant at 5%, and * is significant at 10%. t-statistic is reported in the parenthesis.

Panel A: Hedge funds trading performance

DepVar	**Excess $Ret_{i,s,t+1}$**		**DGTW $Ret_{i,s,t+1}$**	
$DummyPosAHF_{i,s,t} = 1$, if:	$AHF_{i,s,t} > 0$	$AHF_{i,s,t} > 0.5\%$	$AHF_{i,s,t} > 0$	$AHF_{i,s,t} > 0.5\%$
$DummyPosAHF_{i,s,t}$	0.003***	0.004***	0.003***	0.003***
	(7.97)	(7.27)	(7.36)	(7.11)
$SIR_{i,t}$	0.011*	0.012**	0.007	0.007
	(1.99)	(2.14)	(1.40)	(1.54)
$DummyPosAHF_{i,s,t} \times SIR_{i,t}$	0.003***	0.004***	0.003*	0.003***
	(2.82)	(2.99)	(1.98)	(2.95)
Controls	Yes	Yes	Yes	Yes
Year-QTR FE	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
ETF FE	Yes	Yes	Yes	Yes
Observations	922,089	922,089	916,343	916,343
Adjusted R-Squared	0.0793	0.0793	0.0276	0.0276

Panel B: Non-hedge funds trading performance

DepVar	**Excess $Ret_{i,s,t+1}$**		**DGTW $Ret_{i,s,t+1}$**	
$DummyPosAHF_{i,s,t} = 1$, if:	$AHF_{i,s,t} > 0$	$AHF_{i,s,t} > 0.5\%$	$AHF_{i,s,t} > 0$	$AHF_{i,s,t} > 0.5\%$
$DummyPosAHF_{i,s,t}$	-0.002***	-0.002***	-0.002***	-0.002***
	(-3.50)	(-3.68)	(-3.77)	(-3.68)
$SIR_{i,t}$	0.013**	0.013**	0.008	0.008*
	(2.28)	(2.34)	(1.63)	(1.71)
$DummyPosAHF_{i,s,t} \times SIR_{i,t}$	0.000	0.001	0.000	0.001
	(0.04)	(0.41)	(0.07)	(0.26)
Controls	Yes	Yes	Yes	Yes
Year-QTR FE	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
ETF FE	Yes	Yes	Yes	Yes
Observations	919,533	919,533	913,787	913,787
Adjusted R-Squared	0.0793	0.0793	0.0274	0.0274

Table 9: Regress PEAD on hedge funds abnormal holdings and ETFs short interest

Table 9 reports the impact of hedge funds abnormal holdings and ETFs short interest on the PEAD of the stock. The regression model is as follows,

$$CAR(1,k)_{i,s,t+1} = \alpha_i + \alpha_t + \beta_1^C SUERank_{i,s,t+1} \times SIR_{i,t} \times DummyPosAHF_{i,s,t} + \beta_2^C SUERank_{i,s,t+1} \times SIR_{i,t} + \beta_3^C SUERank_{i,s,t+1} \times DummyPosAHF_{i,s,t} + \beta_4^C SIR_{i,t} \times DummyPosAHF_{i,s,t} + \beta_5^C SUERank_{i,s,t+1} + \beta_6^C SIR_{i,t} + \beta_7^C DummyPosAHF_{i,s,t} + controls + \epsilon_{s,t}.$$

CAR is the next quarter's cumulative size-adjusted returns for different post-earnings window. *SUERank* is the quintile ranking of SUE of all stocks in our earnings announcement sample. *DummyPosAHF* is a dummy variable, which equals one if abnormal hedge fund holding on stock *s* is greater than 0% (or 0.5%) in the current quarter. *SIR* is the ETF's short interest ratio at the end of the current quarter. We control for market capitalization, book-to-market ratio, past 12-month returns and interact them with *SUERank*. In addition, we also control for the year-quarter, stock, and ETF fixed effect. All standard errors are clustered by stock and announcement date. We run the above regression model on industry ETFs' member stocks for their quarterly earnings announcements ranging from 1999 to 2017 except for the crisis period (from the fourth quarter of 2006 to the fourth quarter of 2008). *** is significant at 1%, ** is significant at 5%, and * is significant at 10%. t-statistic is reported in the parenthesis.

DepVar	$CAR(1,30)_{i,s,t+1}$	$CAR(1,45)_{i,s,t+1}$	$CAR(1,60)_{i,s,t+1}$	$CAR(1,30)_{i,s,t+1}$	$CAR(1,45)_{i,s,t+1}$	$CAR(1,60)_{i,s,t+1}$
$DummyPosAHF_{i,s,t} = 1$, if:		$AHF_{i,s,t} > 0$			$AHF_{i,s,t} > 0.5\%$	
$SUERank_{i,s,t+1} \times SIR_{i,t} \times DummyPosAHF_{i,s,t}$	-0.007**	-0.015***	-0.022***	-0.014**	-0.020***	-0.029***
	(-1.96)	(-3.32)	(-4.25)	(-2.18)	(-2.75)	(-3.41)
$SUERank_{i,s,t+1} \times SIR_{i,t}$	0.004**	0.010***	0.013***	0.003**	0.007***	0.009***
	(2.05)	(3.50)	(3.93)	(1.97)	(2.97)	(3.18)
$SUERank_{i,s,t+1} \times DummyPosAHF_{i,s,t}$	0.000	0.000	0.000	-0.001	-0.001	-0.002
	(0.40)	(0.50)	(0.34)	(-1.59)	(-0.99)	(-1.51)
$SIR_{i,t} \times DummyPosAHF_{i,s,t}$	0.025**	0.045***	0.065***	0.051***	0.063***	0.090***
	(2.37)	(3.50)	(4.33)	(2.85)	(3.10)	(3.87)
$SUERank_{i,s,t+1}$	0.006***	0.010***	0.011***	0.007***	0.011***	0.012***
	(2.60)	(3.88)	(3.31)	(3.05)	(4.23)	(3.77)
$SIR_{i,t}$	-0.025***	-0.045***	-0.045***	-0.022***	-0.037***	-0.034***
	(-3.38)	(-4.66)	(-3.75)	(-3.55)	(-4.41)	(-3.21)
$DummyPosAHF_{i,s,t}$	0.002	0.002	0.003	0.007***	0.007**	0.011***
	(0.73)	(0.52)	(0.91)	(2.73)	(2.08)	(2.75)
$SUERank_{i,s,t+1} \times Size$	-0.000*	-0.001***	-0.001***	-0.000**	-0.001***	-0.001***
	(-1.80)	(-2.81)	(-2.89)	(-2.02)	(-2.99)	(-3.15)
$SUERank_{i,s,t+1} \times BM$	-0.000	-0.002	-0.001	-0.000	-0.002	-0.001
	(-0.19)	(-1.38)	(-0.52)	(-0.22)	(-1.40)	(-0.56)
$SUERank_{i,s,t+1} \times Mom$	0.001	0.000	0.001	0.001	0.000	0.001

	(0.92)	(0.09)	(0.88)	(0.91)	(0.08)	(0.88)
Size	-0.028***	-0.039***	-0.050***	-0.028***	-0.039***	-0.049***
	(-15.14)	(-16.23)	(-18.34)	(-15.14)	(-16.25)	(-18.29)
BM	-0.006	0.003	-0.002	-0.006	0.003	-0.002
	(-1.20)	(0.47)	(-0.30)	(-1.18)	(0.49)	(-0.27)
Mom	-0.015***	-0.014**	-0.016***	-0.015***	-0.014**	-0.016***
	(-3.38)	(-2.41)	(-2.67)	(-3.36)	(-2.40)	(-2.68)
Year-QTR FE	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
ETF FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	368,827	368,827	368,827	368,827	368,827	368,827
Adjusted R-Squared	0.0715	0.0686	0.0760	0.0717	0.0688	0.0762

Table 10: Short interest on stocks and hedge funds trading on industry ETFs

Table 10 reports results on the correlation between stock-level short interest and ETF-level hedge funds abnormal holdings. The regression model is the following,

$$\Delta SIR(S)_{i,s,t} = \alpha_i + \alpha_t + \beta_1^N DummyNegSUE_{i,s,t+1} + \beta_2^N AHF(E)_{i,t} + \beta_3^N AHF(E)_{i,t} \times DummyNegSUE_{i,s,t+1} + controls + \epsilon_{s,t}.$$

ΔSIR(S) is the current quarter change in the short interest ratio of the stock. *DummyNegSUE* is a dummy and it takes one if SUE of the stock in the coming quarter ranks bottom 25% in our earnings announcement sample. *AHF(E)* is the abnormal holdings by hedge funds of ETF in the current quarter. *AHF(E)* x *DummyNegSUE* is the interaction term. In our control variables, we include market capitalization, book-to-market ratio, institutional ownership, past returns, earnings volatility, and earnings persistence as of the current quarter. In addition, we also control for the year, quarter, and ETF fixed effect. All standard errors are clustered by ETF and year-quarter. We run the above regression model on all underlying stocks of industry ETFs ranging from 1999 to 2017 except for the crisis period (from the fourth quarter of 2006 to the fourth quarter of 2008). *** is significant at 1%, ** is significant at 5%, and * is significant at 10%. t-statistic is reported in the parenthesis.

Dependent: $\boldsymbol{\Delta SIR(S)_{i,s,t}}$		
$AHF(E)_{i,t} \times DummyNegSUE_{i,s,t+1}$	0.005*	0.005**
	(1.87)	(2.05)
$AHF(E)_{i,t}$	0.000	0.003
	(0.01)	(1.09)
$DummyNegSUE_{i,s,t+1}$	0.001***	0.000
	(4.33)	(1.42)
Log(Mktcap)		-0.000**
		(-2.29)
B/M		-0.001
		(-1.40)
Institutional Ownership		0.002
		(1.32)
Reversal		-0.007***
		(-7.76)
Momentum		-0.000
		(-0.12)
Earnings Volatility		-0.000
		(-0.13)
Year FE	Yes	Yes
Quarter FE	Yes	Yes
ETF FE	Yes	Yes
Observations	368,909	355,124
Adjusted R-Squared	0.0090	0.0194

Table 11: Regress PEAD on the industry ETF membership in the match sample

Table 11 reports the regression result on the following model in the match sample,

$$CAR(1,k)_{s,t+1} = \alpha_s + \alpha_t + \beta_1^S SUERank_{s,t+1} + \beta_2^S DummyMember_{s,t} + \beta_3^S SUERank_{s,t+1} \times DummyMember_{s,t} + controls + \epsilon_{s,t}.$$

To form the match sample, in each quarter, we match each industry ETF member stock with a non-member stock from the same industry (Fama and French 12 industries) with smallest sum of absolute difference in the NYSE size and B/M percentile. Panel A and Panel B report pre- and post-matching difference in size and book-to-market ratio between industry ETF member stocks and non-member stocks. Panel C reports regression results in the match sample. *CAR* is the next quarter's cumulative size-adjusted returns for different post-earnings window. *SUERank* is the quintile ranking of SUE of all stocks in our earnings announcement sample. *DummyMember* is a dummy variable, which equals one if the stock is held by at least one industry ETF at the end of the current quarter. We control for market capitalization, book-to-market ratio and interact them with *SUERank*. In addition, we also control for the year-quarter and industry fixed effect. All standard errors are clustered by stock and announcement date. We run the above regression model on stocks in our earnings announcement sample for all quarterly earnings announcement ranging from 1999 to 2017, except for the crisis period (from the fourth quarter of 2006 to the fourth quarter of 2008). *** is significant at 1%, ** is significant at 5%, and * is significant at 10%. t-statistic is reported in the parenthesis.

Panel A: Pre-matching difference in size and book-to-market ratio

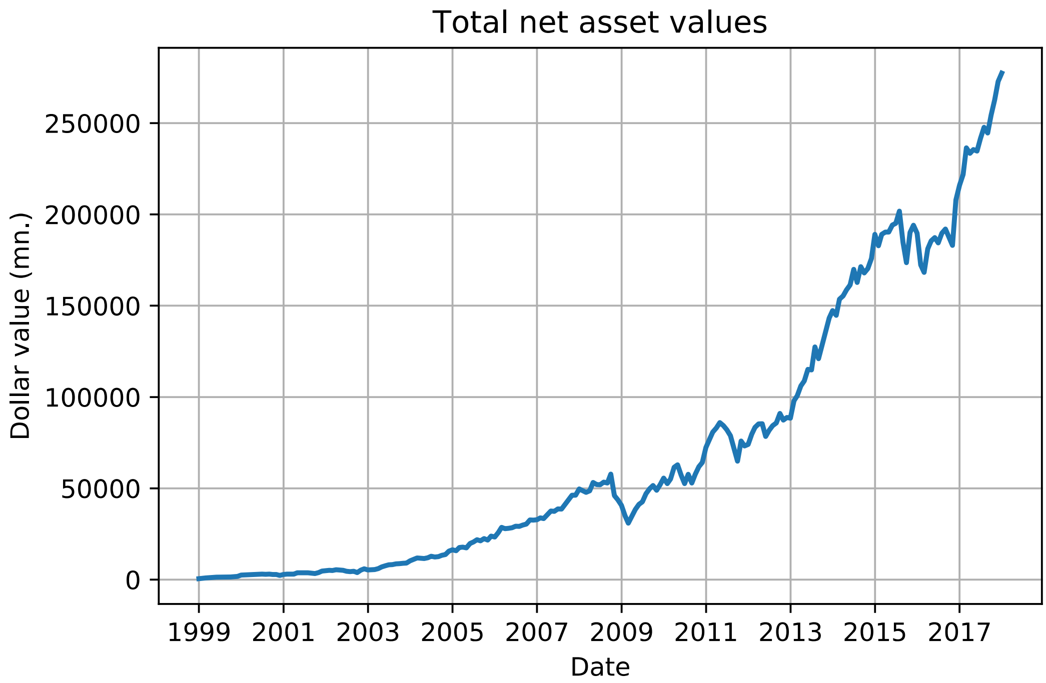
	Member stocks	Non-member stocks	Difference	t-value
# unique firms	2873	8413		
Log(Mktcap)	19.6816	18.6663	-1.0154	-112.45
BM	0.686	0.8235	0.1375	22.94

Panel B: Post-matching difference in size and book-to-market ratio

	Member stocks	Non-member stocks	Difference	t-value
# unique firms	2873	2664		
Log(Mktcap)	19.6816	19.6713	-0.0103	-1.31
BM	0.686	0.6935	0.0074	1.37

Panel C: Regression results in the match sample

DepVar	$\boldsymbol{CAR(1,30)_{s,t+1}}$		$\boldsymbol{CAR(1,45)_{s,t+1}}$		$\boldsymbol{CAR(1,60)_{s,t+1}}$	
$SUERank_{s,t+1}$	0.008***	-0.010	0.011***	0.005	0.012***	-0.006
	(7.05)	(-0.40)	(6.89)	(0.14)	(7.03)	(-0.18)
$DummyMember_{s,t}$	0.020***	0.024***	0.019***	0.023***	0.024***	0.029***
	(4.05)	(4.43)	(3.11)	(3.30)	(3.49)	(3.78)
$SUERank_{s,t+1} \times DummyMember_{s,t}$	-0.004***	-0.005***	-0.005***	-0.005***	-0.006***	-0.007***
	(-2.89)	(-3.06)	(-2.75)	(-2.63)	(-3.15)	(-3.17)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year-QTR FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	38,058	38,058	38,058	38,058	38,058	38,058
Adjusted R-Squared	0.00652	0.00694	0.00649	0.00681	0.00597	0.00626



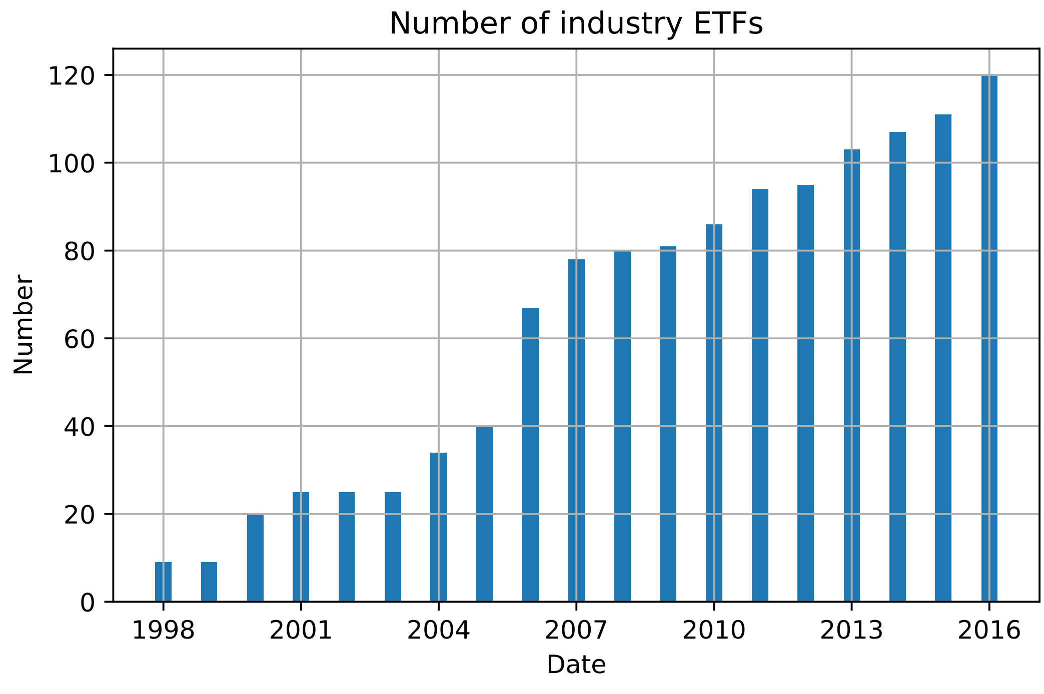


Figure 1: Total net asset values and number of industry ETFs in the sample

Figure 1 shows the time series pattern of total net asset values and number of industry ETFs from 1998.12.16 (the earliest inception date among our industry ETFs) to 2017.12.31 (the end of our sample period).

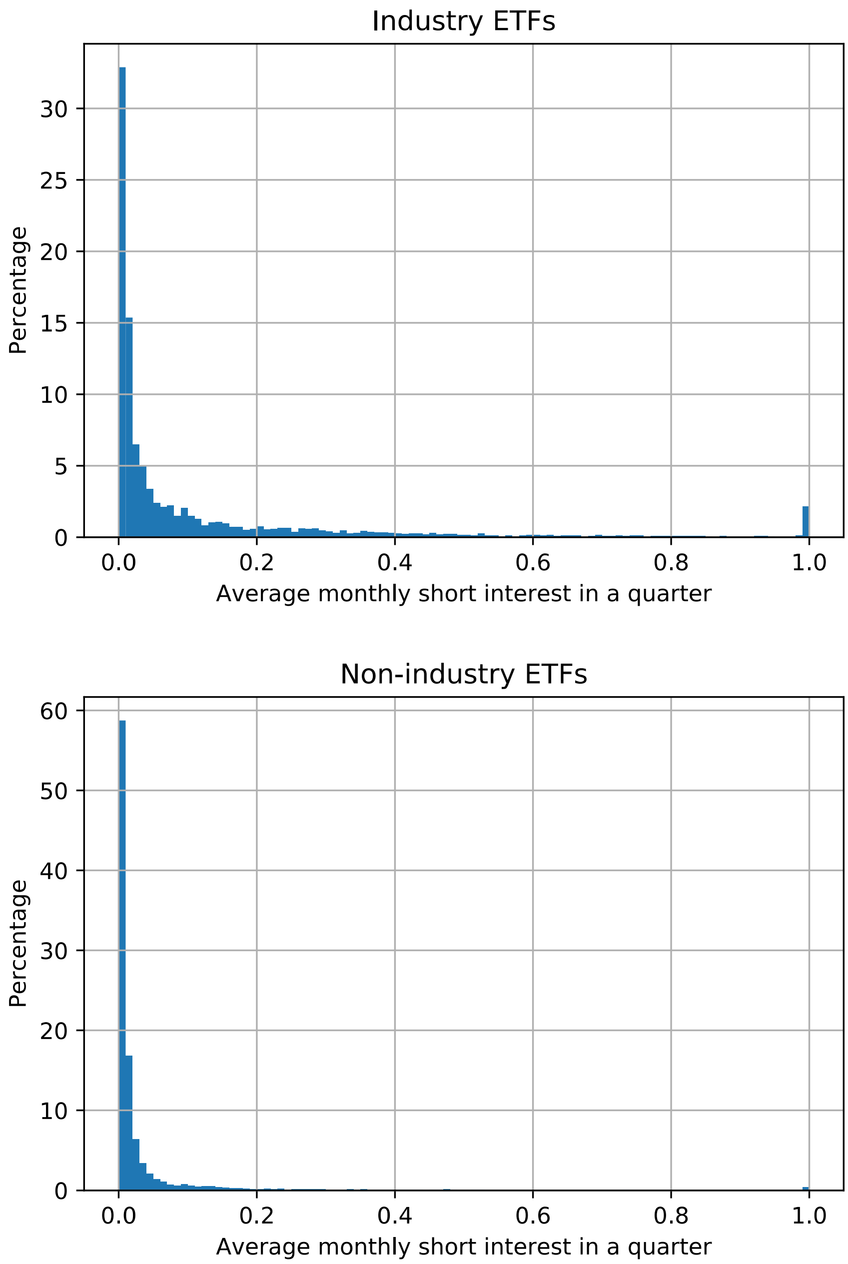


Figure 2: The histogram of the short interest ratio

Figure 2 shows the histogram of the short interest ratio of industry (top panel) and non-industry ETFs (bottom panel), respectively. The short interest ratio is sampled quarterly using the average monthly ratio in a quarter.

Appendix: Additional tests

A1. Additional statistics on short interest

A. The difficulty to short industry ETFs

The high short interest on the industry ETF implies that the ETF cannot be too difficult to short. To verify this conjecture, we study the difficulty to short measure, the *"DCBS"* score in Markit data. A large DCBS indicates that it is more difficult to short. For a particular industry, we compare the value-weighted DCBS for industry ETFs focusing on that industry with the value-weighted DCBS for stocks in that industry. To compute the value-weighted DCBS for industry ETFs, we use the total net asset value to weigh the time series average DCBS for each ETF and obtain the value-weighted average DCBS. For the value-weighted DCBS for stocks, we use the market capitalization to weigh the time series average DCBS for each stock and obtain the value-weighted average DCBS afterwards. We show in Table A.1 that the industry ETF is more accessible to short compared with shorting stocks in an industry.

Figure A.1 shows the 15 most shorted ETFs from Bloomberg, where the top of the list contains mainly industry ETFs. This is consistent with our findings that industry ETFs have more extreme short interest than non-industry ETFs.

[Insert Table A.1 Here]

[Insert Figure A.1 Here]

B. Short interest on stocks

We report short interest on our sample stocks in Table A.2. We split our sample based on the industry ETF membership, and then report short interest for stocks in industry ETFs and not in industry ETFs, respectively.

[Insert Table A.2 here]

A2. Robustness of the main result

A. Robustness tests for Table 3: Alternative definitions of positive SUE

In Table 3, we document that industry ETFs' short interest positively forecasts PosSUE, the likelihood ETFs' constituents reporting good fundamentals. We replicate this result using alternative PosSUE measure.

In Panel A of Table A.3, we use analyst forecast-based SUE (in which SUE is the difference between actual earnings per share and median earnings per share forecast scaled by quarter end price) rather than the seasonal random walk adjusted SUE.

In Panel B, we use a less restrictive industry ETF sample. More specifically, we remove the filter that requires industry ETFs to have at least 30 constituents. With the less restrictive sample, we re-do the analysis in Table 3, and find consistent results.

In Panel C.1 and Panel C.2, we change the cut-off for defining the *"Positive SUE."* Specifically, we define *"Positive SUE"* as top 20% or top 30% *SUE* in our earnings announcement sample. Our results remain unchanged under different definitions of the *"Positive SUE"*.

In Panel D, we conduct a placebo test by replacing the *"Positive SUE"* with *"Negative SUE"* which indicates bottom 25% SUE in our earnings announcement sample. We find that industry ETFs short interest does not predict constituents' likelihood of having *"Negative SUE."*

[Insert Table A.3 Here]

B. Robustness tests for Table 4 and 5: Return predictability of ΔSIR

In Table 4 and Table 5, we document that ΔSIR positively predicts industry ETF returns. We replicate these results using NAV change as an alternative measure for industry ETF returns. Table A.4 shows that ΔSIR also positively predicts NAV change (Panel A and Panel B), which is consistent with our results on return predictability. Further, we report, in Panel C of Table A.4, the monthly performance of industry ETF long-short portfolios (sorted on ΔSIR) for 12 months after the formation. We find that positive significant long-short returns appear in the first month after formation and returns do not reverse afterward.

[Insert Table A.4 Here]

C. Rule out alternative explanations for return predictability of ΔSIR

To rule out the possibility that high ΔSIR predicts positive industry ETF price premiums (or negative price discounts) instead of high returns in the future, we run the predictive regression of industry ETF discount on ΔSIR. We find no significant relationship between industry ETF discount and ΔSIR (see Panel A of Table A.5).

[Insert Table A.5 Here]

Brown, Davies and Ringgenberg (2018) find that arbitrage activity on ETFs, measured by change in shares outstanding, negatively predicts subsequent returns among broad equity ETFs, but not among industry ETFs. We examine this result in our industry ETF sample.

Panel B of Table A.5 reports no return predictability of change in shares outstanding on returns in the subsequent month, which is consistent with the previous study.

A3. Crisis period versus non-crisis period

A. Return predictability of ΔSIR in the crisis and non-crisis period

We compare the performance of industry ETF long-short portfolios (sorted on ΔSIR) in the crisis period with their performance in the non-crisis period. The crisis period is defined as the period from the fourth quarter of 2006 to the last quarter of 2008. We find that the high ΔSIR industry ETF portfolio performs worse than the low ΔSIR portfolio during the crisis period. Returns of the ΔSIR long-short portfolio in the crisis period are significantly lower than that in the non-crisis period, with signs flipped.

[Insert Table A.6 Here]

Table A.7 compares performances of the ΔSIR long-short portfolios on industry ETFs' member stocks between the crisis period and the non-crisis period.

[Insert Table A.7 Here]

B. Hedge fund trading and industry ETF short interest during the crisis

We document in Table 6 that, in the non-crisis period, hedge funds abnormal holdings on stocks is positively correlated with contemporaneous high short interest on industry ETFs, especially before positive SUE. In Table A.8, we re-evaluate this result in the crisis period and find no significant contemporaneous correlation between hedge funds abnormal holdings and industry ETFs short interest, neither conditionally nor unconditionally.

[Insert Table A.8 here]

A4. Additional results on industry ETFs short interest and informed trading

A. Source of hedge funds trading performance

In Table 8, we show the contemporaneous correlation between hedge funds trading and industry ETFs short interest positively predicts stock returns. The predictability implies that hedge funds have high trading performance. We examine the source of this high trading performance in Table A.9. We find that contemporaneous high abnormal hedge funds holdings and industry ETF short interest predicts higher firm-specific fundamentals, such as positive earnings, higher ROA, and higher ROE.

[Insert Table A.9 Here]

B. Short sellers trading performance

Informed short sellers can long industry ETFs to hedge their short positions in particular stocks. We have shown this *"long-the-ETF/short-the-stock"* in Table 10. In Table A.10, we further examine this strategy by studying short sellers trading performance. We first show that, before the release of negative earnings, high stock-level short interest ratio is positively correlated with high abnormal hedge funds holdings on industry ETFs (Panel A). This positive significant correlation only exists among large stocks, which is consistent with small stocks being difficult to short. Then we find that an increase in stock-level short selling and high abnormal hedge funds holdings on industry ETFs indeed predicts lower subsequent stock returns (Panel B).

[Insert Table A.10 Here]

A5. Additional results on industry ETF membership and market efficiency

A. Industry ETF membership and PEAD in the full sample

In Table 11, we find that industry ETF membership reduces the PEAD in the propensity score matched sample. In Table A.11, we show the effect of the industry ETF membership in the full sample. We find that results are consistent, but with stronger significance (in terms of larger *t-stat.*).

[Insert Table A.11 Here]

Further, we conduct a placebo test to examine the effect of the non-industry ETF membership on the PEAD. Table A.12 shows that the non-industry ETF membership has no significant influence on the PEAD.

[Insert Table A.12 Here]

B. Industry ETF membership and other marker efficiency measures

We report the effect of industry ETF membership on other price efficiency measures, including the *"price delay"* and *"variance ratio"* in Table A.13. We find consistent evidence that the industry ETF membership improves market efficiency.

[Insert Table A.13 Here]

15 Most Shorted ETFs (excluding inverse/leveraged)

Ticker	Fund	Short Interest %
XRT	SPDR S&P Retail ETF	507.99
SMH	VanEck Vectors Semiconductor ETF	176.58
UNG	United States Natural Gas Fund LP	106.57
XOP	SPDR S&P Oil & Gas Exploration & Production ETF	89.41
VXX	iPath S&P 500 VIX Short-Term Futures ETN	80.95
FXE	CurrencyShares Euro Trust	78.56
XBI	SPDR S&P BIOTECH ETF	59.43
IMED	PureFunds ETFx HealthTech ETF	58.63
DWLV	PowerShares DWA Momentum & Low Volatility Rotation Portfolio	57.34
FXY	CurrencyShares Japanese Yen Trust	55.53
CTNN	iPath Pure Beta Cotton ETN	50.00
EWW	iShares MSCI Mexico Capped ETF	48.07
IYR	iShares U.S. Real Estate ETF	47.92
ERY	Direxion Daily Energy Bear 3X Shares	46.39
OIH	VanEck Vectors Oil Services ETF	42.75

Source: Bloomberg; data as of July 17, 2017

Figure A.1: The 15 mostly shorted ETFs from Bloomberg

Figure A.1 is the rank of 15 mostly shorted ETFs based on Bloomberg data.

Table A.1: Industry ETFs and shorting difficulty

Table A.1 reports the difficulty, the DCBS score from Markit, to short industry ETFs. To compute the value-weighted DCBS for industry ETFs, we use the total net asset value to weigh the time series average DCBS for each ETF and obtain the value-weighted average DCBS. For the value-weighted DCBS for stocks, we use the market capitalization to weigh the time series average DCBS for each stock and obtain the value-weighted average DCBS afterwards.

FF12 industry classification	Value-weighted DCBS	
	Industry ETFs	Stocks
1	0.984	4.153
2	N.A.	4.043
3	1.574	3.932
4	0.877	4.136
5	1.645	3.687
6	1.781	4.333
7	N.A.	7.732
8	0.930	4.320
9	1.434	4.415
10	1.271	4.529
11	0.972	4.774
12	4.185	9.162

Table A.2: Stock level short interest

Table A.2 reports the summary statistics of stock level short interest for our sample stocks. We break our sample stocks into two groups: 1). Stocks with industry ETF membership; 2). Stocks without industry ETF membership.

	Mean	**Std.**	**5%**	**25%**	**50%**	**75%**	**95%**
In industry ETFs	0.052	0.058	0.003	0.015	0.032	0.068	0.167
Out industry ETFs	0.040	0.054	0.0	0.006	0.022	0.052	0.146

Table A3: Robustness of Table 3

Table A.3 reports robustness checks for results in Table 3. In Panel A, we replicate Table 3 using the analyst forecast based SUE definition, which is the difference between actual EPS and median EPS forecast value scaled by quarter-end stock price. In Panel B, we include industry ETFs with number of holding stocks smaller than 30 and replicate Table 3. In Panel C.1 and Panel C.2, we replicate Table3 using different cut-offs for defining positive SUE. In Panel D, we define negative earnings ratio for a ETF in a quarter as number of holding stocks with negative SUE in the quarter divided by total number of holding stocks at the quarter-end. We define bottom 25% SUE in our sample as negative SUE. We run similar regressions as those in Table 3, replacing positive earnings ratio by negative earnings ratio.

Panel A: Replicate Table 3 using analyst forecast based SUE

DepVar: $PosSUE_{i,t}$	**Industry ETFs**	**Non-industry ETFs**	**All ETFs**
$SIR_{i,t-1}$	0.024*	-0.025**	-0.028***
	(1.92)	(-2.52)	(-3.18)
$SIR_{i,t-1} \times DummyIndetf_i$			0.051***
			(4.11)
$SIR_{i,t-1} \times DummyCrisis_t$	-0.047***	-0.014	-0.013
	(-3.40)	(-1.03)	(-1.04)
$SIR_{i,t-1} \times DummyIndetf_i \times DummyCrisis_t$			-0.033**
			(-2.11)
$DummyCrisis_t$	0.028**	0.026***	0.027***
	(2.57)	(2.79)	(2.77)
$DummyCrisis_t \times DummyIndetf_i$			-0.002
			(-0.39)
Year FE	Yes	Yes	Yes
Qtr FE	Yes	Yes	Yes
ETF FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	4,778	9,638	14,416
Adjusted R-Squared	0.535	0.735	0.655

Panel B: Replicate Table 3 in all industry ETF sample

DepVar: $PosSUE_{i,t}$	**Industry ETFs**	**Non-industry ETFs**	**All ETFs**
$SIR_{i,t-1}$	0.042**	-0.020	-0.024*
	(2.55)	(-1.40)	(-1.68)
$SIR_{i,t-1} \times DummyIndetf_i$			0.064***
			(3.67)
$SIR_{i,t-1} \times DummyCrisis_t$	-0.083***	-0.003	-0.002
	(-6.35)	(-0.21)	(-0.12)
$SIR_{i,t-1} \times DummyIndetf_i \times DummyCrisis_t$			-0.081***
			(-5.09)
$DummyCrisis_t$	-0.005	-0.068**	-0.043*
	(-0.24)	(-2.39)	(-1.71)
$DummyCrisis_t \times DummyIndetf_i$			0.008
			(1.49)
Year FE	Yes	Yes	Yes
Qtr FE	Yes	Yes	Yes
ETF FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	5,431	9,638	15,069
Adjusted R-Squared	0.501	0.683	0.585

Panel C.1: Replicate Table 3. Define top 20% SUE as Positive SUE

DepVar: $PosSUE_{i,t}$	**Industry ETFs**	**Non-industry ETFs**	**All ETFs**
$SIR_{i,t-1}$	0.046***	-0.019	-0.021
	(2.78)	(-1.40)	(-1.64)
$SIR_{i,t-1} \times DummyIndetf_i$			0.066***
			(4.00)
$SIR_{i,t-1} \times DummyCrisis_t$	-0.054***	0.008	0.009
	(-3.70)	(0.62)	(0.75)
$SIR_{i,t-1} \times DummyIndetf_i \times DummyCrisis_t$			-0.062***
			(-3.90)
$DummyCrisis_t$	-0.014	-0.058**	-0.039
	(-0.60)	(-2.14)	(-1.57)
$DummyCrisis_t \times DummyIndetf_i$			-0.001
			(-0.10)
Year FE	Yes	Yes	Yes
Qtr FE	Yes	Yes	Yes
ETF FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	4,778	9,638	14,416
Adjusted R-Squared	0.519	0.682	0.598

Panel C.2: Replicate Table 3. Define top 30% SUE as Positive SUE

DepVar: $PosSUE_{i,t}$	**Industry ETFs**	**Non-industry ETFs**	**All ETFs**
$SIR_{i,t-1}$	0.055***	-0.025*	-0.023
	(2.79)	(-1.68)	(-1.66)
$SIR_{i,t-1} \times DummyIndetf_i$			0.070***
			(3.56)
$SIR_{i,t-1} \times DummyCrisis_t$	-0.083***	-0.006	-0.007
	(-5.09)	(-0.34)	(-0.45)
$SIR_{i,t-1} \times DummyIndetf_i \times DummyCrisis_t$			-0.075***
			(-3.88)
$DummyCrisis_t$	-0.007	-0.077**	-0.051**
	(-0.34)	(-2.62)	(-2.02)
$DummyCrisis_t \times DummyIndetf_i$			0.008
			(1.41)
Year FE	Yes	Yes	Yes
Qtr FE	Yes	Yes	Yes
ETF FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	4,778	9,638	14,416
Adjusted R-Squared	0.484	0.676	0.580

Panel D: Forecast negative earnings ratio (*NegSUE*)

DepVar: $NegSUE_{i,t}$	**Industry ETFs**	**Non-industry ETFs**	**All ETFs**
$SIR_{i,t-1}$	0.004	-0.010	-0.022*
	(0.17)	(-0.75)	(-1.73)
$SIR_{i,t-1} \times DummyIndetf_i$			0.047**
			(2.38)
$SIR_{i,t-1} \times DummyCrisis_t$	0.067***	0.004	0.010
	(2.97)	(0.27)	(0.76)
$SIR_{i,t-1} \times DummyIndetf_i \times DummyCrisis_t$			0.057**
			(2.03)
$DummyCrisis_t$	0.063***	0.144***	0.115***
	(2.88)	(4.20)	(4.06)
$DummyCrisis_t \times DummyIndetf_i$			-0.017***
			(-2.90)
Year FE	Yes	Yes	Yes
Qtr FE	Yes	Yes	Yes
ETF FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	4,778	9,638	14,416
Adjusted R-Squared	0.574	0.711	0.641

Table A.4: Robustness check for Table 4 and 5

Table A.4 reports robustness checks for results in Table 4 and Table 5. In Panel A, we replicate Fama-MacBeth regression results in Table 4 using NAV change as dependent variable. In our control variables, we include stock characteristics as of month *t* end, including past 12-month returns, market capitalization, book-to-market ratio, asset growth, operating profitability, gross profitability, investment growth, net issuance, accruals, and net operating assets. In Panel B, we measure industry ETF returns by NAV change and report performance of industry ETF portfolios sorted on ΔSIR. In Panel C, we form industry ETF portfolios sorted on ΔSIR (bottom 30% / middle 40% / top 30%) at the end of month t and track their performance from month t+1 to month t+12. To deal with overlapping portfolios in holding period, we take equally-weighted returns of portfolios formed in different months as holding period returns. We report the long-short portfolio returns (long top 30% and short bottom 30%) in different holding periods.

Panel A: Fama-MacBeth Regression of NAV Change on ΔSIR

	DepVar: $NAV\ Change_{t+1}$	
ΔSIR_t	0.034***	0.026**
	(3.13)	(2.57)
Intercept	0.012***	0.013
	(3.60)	(1.16)
Controls	Yes	Yes

Panel B: NAV Change of industry ETF portfolios sorted on ΔSIR

	Excess Returns		CAPM alpha		3-factor alpha		4-factor alpha	
Portfolio	coef.	t-stat	coef.	t-stat	coef.	t-stat	coef.	t-stat
Bottom 30%	0.97	2.75	-0.45	-6.42	-0.40	-6.18	-0.39	-6.29
Mid 40%	1.16	3.45	-0.21	-2.83	-0.17	-2.59	-0.16	-2.49
Top 30%	1.25	3.59	-0.17	-1.79	-0.12	-1.35	-0.11	-1.26
Top - Bottom	**0.29**	**3.31**	**0.29**	**3.20**	**0.28**	**3.15**	**0.28**	**3.07**

Panel C: Industry ETF long-short portfolios sorted on ΔSIR

	Excess Returns		CAPM alpha		3-factor alpha		4-factor alpha	
Holding Period	coef.	t-stat	coef.	t-stat	coef.	t-stat	coef.	t-stat
Month t+1	**0.24**	**2.65**	**0.24**	**2.59**	**0.23**	**2.50**	**0.23**	**2.40**
Month t+2 to Month t+12	0.03	1.22	0.02	0.75	0.02	0.83	0.03	0.98

Table A.5: Alternative explanations for short interest return predictability

Table A.5 examines alternative explanations for return predictability of ΔSIR. Panel A reports Fama-MacBeth regression of industry ETF discount in the next month on change in the short interest ratio of the current month. We compute discount as one minus the ratio of month-end share price divided by month-end NAV. In our control variables, we include stock characteristics as of month *t* end, including past 12-month returns, market capitalization, book-to-market ratio, asset growth, operating profitability, gross profitability, investment growth, net issuance, accruals, and net operating assets. Panel B reports Fama-MacBeth regression of industry ETF returns in the next month on change in shares outstanding (ΔShrout) and change in short interest ratio in the current month. The sample period of our analysis is from January 2005 to December 2017, excluding crisis period (from October 2006 to December 2008).

Panel A: Fama-MacBeth regression of industry ETF discount on ΔSIR

DepVar: $\boldsymbol{Discount_{t+1}}$	**Discount = 1-PRC/NAV**		**Discount=abs(1-PRC/NAV)**	
ΔSIR_t	0.000	-0.000	0.000	0.001
	(0.31)	(-0.80)	(1.39)	(1.21)
Intercept	-0.000	-0.001	0.001***	0.003***
	(-3.49)	(-1.47)	(21.77)	(11.16)
Controls	Yes	Yes	Yes	Yes

Panel B: Fama-MacBeth regression of industry ETF return on ΔShrout

DepVar: $\boldsymbol{Ret_{t+1}}$		
$\Delta Shrout_t$	-0.003	-0.002
	(-0.54)	(-0.65)
ΔSIR_t	0.031***	0.022**
	(2.87)	(2.30)
Intercept	0.013***	0.020
	(3.74)	(1.70)
Controls	Yes	Yes

Table A.6: Industry ETF portfolio sorting on ΔSIR in the crisis and non-crisis period

Table A.6 compares performance of industry ETF portfolios sorted on ΔSIR in the crisis and non-crisis period. At the end of each month, we sort industry ETFs into three groups (bottom 30%/ mid40% / top 30%) based on ΔSIR in that month. We hold the portfolios in the next month and compute equally-weighted portfolio returns. Panel A reports holding period returns from January 2005 to December 2017, excluding crisis period (from October 2006 to December 2008). Panel B reports holding period returns in the crisis period (from October 2006 to December 2008). Panel C compares the difference in long-short portfolio returns (long top 30% and short bottom 30%) between the crisis and non-crisis period.

Panel A: Non-crisis period

	Excess Returns		CAPM alpha		3-factor alpha		4-factor alpha	
Portfolio	coef.	t-stat	coef.	t-stat	coef.	t-stat	coef.	t-stat
Bottom 30%	1.13	3.07	-0.29	-4.31	-0.23	-3.94	-0.21	-3.84
Mid 40%	1.29	3.72	-0.08	-1.14	-0.04	-0.58	-0.02	-0.41
Top 30%	1.37	3.78	-0.05	-0.54	0.00	0.01	0.01	0.17
Top – Bottom	**0.24**	**2.65**	**0.24**	**2.59**	**0.23**	**2.50**	**0.23**	**2.40**

Panel B: Crisis period

	Excess Returns		CAPM alpha		3-factor alpha		4-factor alpha	
Portfolio	coef.	t-stat	coef.	t-stat	coef.	t-stat	coef.	t-stat
Bottom 30%	-1.68	-1.39	0.19	0.76	0.08	0.32	0.16	0.57
Mid 40%	-1.61	-1.23	0.48	2.54	0.33	2.35	0.31	2.33
Top 30%	-2.09	-1.65	-0.03	-0.18	-0.02	-0.10	-0.07	-0.43
Top – Bottom	**-0.42**	**-1.23**	**-0.22**	**-0.77**	**-0.10**	**-0.34**	**-0.23**	**-0.71**

Panel C: Difference in long-short returns

	Excess Returns		CAPM alpha		3-factor alpha		4-factor alpha	
Portfolio	coef.	t-stat	coef.	t-stat	coef.	t-stat	coef.	t-stat
Difference in Top – Bottom	**-0.66**	**-2.43**	**-0.52**	**-1.92**	**-0.47**	**-1.78**	**-0.49**	**-1.87**

Table A.7: Member stocks portfolio sorting on ΔSIR in the crisis and non-crisis period

Table A.7 compares performance of portfolios of industry ETF member stock sorting on ΔSIR in the crisis and non-crisis period. At the end of each month, we sort member stocks into deciles based on ΔSIR in the month. We hold the portfolios in the next month and compute equally-weighted portfolio returns. Panel A reports holding period returns from January 1999 to December 2017, excluding the crisis period (from October 2006 to December 2008). Panel B reports holding period returns in the crisis period: from October 2006 to December 2008. Panel C compares the difference in long-short portfolio returns between crisis and non-crisis period.

Panel A: Non-Crisis period

	Excess Returns		CAPM alpha		3-factor alpha		4-factor alpha	
Decile	coef.	t-stat	coef.	t-stat	coef.	t-stat	coef.	t-stat
1 (low)	1.25	3.15	0.36	1.47	0.01	0.09	0.06	0.38
10 (high)	0.85	1.97	-0.12	-0.50	-0.46	-2.46	-0.39	-2.16
10-1	**-0.40**	**-3.22**	**-0.47**	**-3.97**	**-0.47**	**-3.93**	**-0.45**	**-3.67**

Panel B: Crisis period

	Excess Returns		CAPM alpha		3-factor alpha		4-factor alpha	
Decile	coef.	t-stat	coef.	t-stat	coef.	t-stat	coef.	t-stat
1 (low)	-2.45	-1.70	-0.14	-0.36	-0.26	-1.11	-0.15	-0.57
10 (high)	-2.70	-1.99	-0.43	-0.86	-0.44	-1.39	-0.18	-0.68
10-1	**-0.25**	**-0.62**	**-0.28**	**-0.87**	**-0.17**	**-0.52**	**-0.02**	**-0.07**

Panel C: Difference in long-short returns

	Excess Returns		CAPM alpha		3-factor alpha		4-factor alpha	
	coef.	t-stat	coef.	t-stat	coef.	t-stat	coef.	t-stat
Difference	**0.15**	**0.42**	**0.33**	**0.93**	**0.31**	**0.86**	**0.30**	**0.84**

Table A.8: Hedge funds trading and ETF short interest in the crisis period

Table A.8 examines the relationship between abnormal hedge fund holding and ETF short interest ratio in the financial crisis period (from the fourth quarter of 2006 to the fourth quarter of 2008). The regression specification and variable definitions are the same with Table 6,

$$AHF_{i,s,t} = \alpha_i + \alpha_t + \beta_1^H DummyPosSUE_{i,s,t+1} + \beta_2^H SIR_{i,t} + \beta_3^H SIR_{i,t} \times DummyPosSUE_{i,s,t+1} + controls + \epsilon_{s,t}.$$

*** is significant at 1%, ** is significant at 5%, and * is significant at 10%. T-statistic is reported in the parenthesis.

DepVar: $AHF_{i,s,t}$		
$DummyPosSUE_{i,s,t+1}$	0.009	-0.073
	(0.11)	(-1.35)
$SIR_{i,t}$	-0.000	-0.009
	(-0.00)	(-0.11)
$SIR_{i,t} \times DummyPosSUE_{i,s,t+1}$	0.009	0.005
	(0.11)	(0.06)
Log(Mktcap)		-0.033
		(-1.47)
BM		-0.051
		(-1.43)
Institutional Ownership		0.254
		(1.19)
Reversal		-0.106
		(-0.78)
Momentum		-0.392***
		(-3.79)
Earnings Volatility		0.001
		(0.38)
Earnings Persistence		0.092
		(1.22)
Year FE	Yes	Yes
Quarter FE	Yes	Yes
ETF FE	Yes	Yes
Observations	63,731	62,009
Adjusted R-Squared	0.0297	0.0351

Table A.9: Forecast fundamentals by hedge funds trading and ETF short interest

Table A.9 reports the regression result on the following model,

$$Fundamentals_{i,s,t+1},$$
$$= \alpha_i + \alpha_t + \beta_1^F AHF_{i,s,t} + \beta_2^F SIR_{i,t} + \beta_3^F AHF_{i,s,t} \times SIR_{i,t} + controls$$
$$+ \epsilon_{s,t}.$$

The dependent variables, *Fundamentals*, include *DummyPosSUE*, *ROA*, *Ind-adj ROA*, *ROE*, and *Ind-adj ROE*. *DummyPosSUE* is a dummy variable, which equals one if standardized unexpected earnings ranks top 25% in our sample. We compute quarterly ROA as income before extraordinary items divided by total assets. Industry-adjusted ROA is firm ROA minus value-weighted industry average ROA. We compute ROE as income before extraordinary items divided by book equity, and we compute industry adjusted ROE as firm ROE minus value-weighted industry average ROE. The dependent variable *ROA* (*Ind-adj ROA*, *ROE*, or *Ind-adj ROE)* is the average quarterly *ROA* (*Ind-adj ROA*, *ROE*, or *Ind-adj ROE)* from quarter *t+1* to quarter *t+4*. *AHF* is the abnormal holdings by hedge funds of the stock in the current quarter standardized by the stock's total shares outstanding at the end of the current quarter. *SIR* is the end of quarter short interest ratio on the ETF. In our control variables, we include market capitalization, book-to-market ratio, institutional ownership, past returns, earnings volatility, and earnings persistence as of quarter *t* end. In addition, we also control for the year, quarter, and ETF fixed effect. The sample period of our analysis is from January 2005 to December 2017, excluding crisis period (from October 2006 to December 2008). Standard errors are clustered by ETF and year-quarter. *** is significant at 1%, ** is significant at 5%, and * is significant at 10%. T-statistic is reported in the parenthesis.

DepVar:	$\boldsymbol{DummyPosSUE_{i,s,t+1}}$	$\boldsymbol{ROA_{i,s,t+1}}$	***Ind Adj*** $\boldsymbol{ROA_{i,s,t+1}}$	$\boldsymbol{ROE_{i,s,t+1}}$	***Ind Adj*** $\boldsymbol{ROE_{i,s,t+1}}$
$AHF_{i,s,t}$	0.137	-0.020***	-0.019***	-0.060***	-0.054***
	(1.56)	(-4.40)	(-4.17)	(-3.28)	(-3.05)
$SIR_{i,t}$	0.018	-0.000	0.001	-0.001	0.002
	(0.77)	(-0.23)	(1.20)	(-0.23)	(0.42)
$AHF_{i,s,t} \times SIR_{i,t}$	0.590***	0.031**	0.030**	0.073**	0.060**
	(4.06)	(2.64)	(2.56)	(2.28)	(2.12)
Log(Mktcap)	-0.022***	0.006***	0.006***	0.011***	0.011***
	(-7.59)	(6.50)	(7.14)	(4.99)	(5.34)
BM	0.074***	-0.016***	-0.014***	-0.042***	-0.037***
	(5.29)	(-7.30)	(-7.47)	(-10.24)	(-9.56)
Institutional Ownership	-0.011	0.004***	0.003***	0.010***	0.010***
	(-0.81)	(3.64)	(2.87)	(3.05)	(2.81)
Reversal	0.012	0.009***	0.007***	0.023***	0.021***
	(0.47)	(3.13)	(3.07)	(3.88)	(3.87)
Momentum	0.084***	0.007***	0.006***	0.015***	0.014***
	(4.95)	(5.59)	(5.31)	(5.79)	(5.49)
Earnings Volatility	0.001**	0.003	0.004	-0.026**	-0.023**
	(2.05)	(0.70)	(1.03)	(-2.29)	(-2.03)
Earnings Persistence	-0.041***	-0.001***	-0.001**	-0.003**	-0.002*
	(-4.17)	(-2.78)	(-2.39)	(-2.45)	(-1.83)
Year FE	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes
ETF FE	Yes	Yes	Yes	Yes	Yes
Observations	358,045	355,227	355,227	355,227	355,227
Adjusted R-Squared	0.065	0.750	0.754	0.551	0.542

Table A.10: Trading performance of short sellers

Table A.10 evaluates trading performance of short sellers. We cut our sample into smaller 50% firms and larger 50% firms in each quarter. In Panel A, we report the quarterly regression result on the following model in smaller and larger firm sample respectively,

$$\Delta SIR(S)_{i,s,t}, = \alpha_i + \alpha_t + \beta_1^{SS} DummyNegSUE_{i,s,t+1} + \beta_2^{SS} AHF(E)_{i,t} + \beta_3^{SS} AHF_{i,t} \times DummyNegSUE_{i,s,t+1} + controls + \epsilon_{s,t}.$$

ΔSIR(S) is the stock-level change in short interest ratio in the current quarter. *DummyNegSUE* is a dummy and it takes one if the stock SUE in the next quarter ranks bottom 25% in our sample. *AHF(E)* is the ETF-level abnormal holdings by hedge funds in the current quarter. In panel B, we focus on the sample of larger 50% stocks, and report the monthly regression results on following model,

$$Ret_{i,s,t+1}, = \alpha_i + \alpha_t + \theta_1^{SS} AHF(E)_{i,t} \times DummyPos\Delta SIR(S)_{i,s,t} + \theta_2^{SS} AHF(E)_{i,t} \times DummyPos\Delta SIR(S)_{i,s,t} \times Large_{i,s,t} + \theta_3^{SS} AHF(E)_{i,t} + \theta_4^{SS} AHF(E)_{i,t} \times Large_{i,s,t} + \theta_5^{SS} DummyPos\Delta SIR_{i,s,t} + \theta_6^{SS} DummyPos\Delta SIR_{i,s,t} \times Large_{i,s,t} + controls + \epsilon_{s,t}.$$

Ret is stock returns (excess returns or DGTW characteristics-adjusted returns) in the following month. *DummyPos ΔSIR* is a dummy variable, which equals one if *ΔSIR* is greater than 0 (or 0.5%) in the current month. $Large_{i,s,t}$ is a dummy variable, which equals one if the stocks is among the larger 50% stocks in our sample at the end of the current month. All standard errors are clustered by ETF and year-quarter. The sample period of our analysis in Panel A and Panel B is from January 2005 to December 2017, excluding crisis period (from October 2006 to December 2008). Standard errors are clustered by ETF and year-quarter. *** is significant at 1%, ** is significant at 5%, and * is significant at 10%. t-statistic is reported in the parenthesis.

Panel A: Stock ΔSIR and ETF AHF

DepVar: $\Delta SIR_{i,s,t}$	***Larger Firms***		***Smaller Firms***	
$AHF_{i,t} \times DummyNegSUE_{i,s,t+1}$	0.006*	0.006*	0.002	0.003
	(1.86)	(1.72)	(0.55)	(0.71)
$AHF_{i,t}$	-0.003	0.002	0.005	0.007
	(-1.36)	(0.91)	(0.96)	(1.41)
$DummyNegSUE_{i,s,t+1}$	0.001*	-0.000	0.001***	0.001**
	(1.84)	(-0.78)	(3.33)	(2.28)
Log(Mktcap)		-0.000		0.000
		(-1.62)		(1.27)
BM		0.000		-0.001*
		(0.30)		(-1.75)
Institutional Ownership		0.003		0.001
		(1.30)		(0.98)
Reversal		-0.012***		-0.004***
		(-12.62)		(-4.76)
Momentum		0.000		0.000
		(0.06)		(0.21)
Earnings Volatility		-0.000		0.000
		(-0.67)		(0.57)
Year FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
ETF FE	Yes	Yes	Yes	Yes
Observations	187,033	179,703	181,875	175,420
Adjusted R-Squared	0.0055	0.0368	0.0167	0.0206

Panel B: Predict Stock Returns by lagged Stock ΔSIR and ETF AHF

DepVar:	***Excess*** $\boldsymbol{Ret_{i,s,t+1}}$		***DGTW*** $\boldsymbol{Ret_{i,s,t+1}}$	
$DummyPos\Delta SIR(S)_{i,s,t}$ = 1 if:	$\Delta SIR_{i,s,t} > 0$	$\Delta SIR_{i,s,t} > 0.5\%$	$\Delta SIR_{i,s,t} > 0$	$\Delta SIR_{i,s,t} > 0.5\%$
$AHF_{i,t} \times DummyPos\Delta SIR_{i,s,t}$	0.034	0.068	0.045**	0.068**
	(0.92)	(1.40)	(2.15)	(2.62)
$AHF_{i,t} \times DummyPos\Delta SIR_{i,s,t} \times Large_{i,s,t}$	-0.085*	-0.154**	-0.054*	-0.089**
	(-1.85)	(-2.05)	(-1.97)	(-2.22)
$AHF_{i,t}$	-0.014	-0.012	-0.022*	-0.015**
	(-1.02)	(-1.43)	(-1.79)	(-2.24)
$AHF_{i,t} \times Large_{i,s,t}$	0.033*	0.016	0.030*	0.020**
	(1.87)	(1.34)	(1.95)	(2.31)
$Pos\_\Delta SIR_{i,s,t}$	0.003	0.002	-0.000	-0.001
	(1.05)	(0.83)	(-0.29)	(-1.11)
$Pos\_\Delta SIR_{i,s,t} \times Large_{i,s,t}$	-0.000	0.001	-0.002	-0.000
	(-0.08)	(0.54)	(-1.60)	(-0.14)
$Large_{i,s,t}$	-0.001	-0.001	0.002	0.001
	(-0.33)	(-0.58)	(1.52)	(0.90)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
ETF FE	Yes	Yes	Yes	Yes
Year-QTR FE	Yes	Yes	Yes	Yes
Observations	1,209,253	1,209,253	1,206,035	1,206,035
Adjusted R-Squared	0.0978	0.0977	0.0468	0.0467

Table A.11: Regress PEAD on industry ETF membership

Table A.11 reports the regression result on the following model

$$CAR(1,k)_{s,t+1} = \alpha_s + \alpha_t + \beta_1^S SUERank_{s,t+1} + \beta_2^S DummyMember_{s,t} + \beta_3^S SUERank_{s,t+1} \times DummyMember_{s,t} + controls + \epsilon_{s,t}.$$

CAR is cumulative size-adjusted returns for different post-earnings-announcement window. *SUERank* is the quintile ranking of SUE in our sample. *DummyMember* is a dummy variable, which equals one if the stock is held by at least one industry ETF in the current earnings announcement quarter. We control for market capitalization, book-to-market ratio, past 12-month returns and interact them with *SUERank*. In addition, we also control for the year-quarter, firm, and ETF fixed effect. All standard errors are clustered by stock and announcement date.. We run the above regression model on our earnings announcements ranging from 1999 to 2017 except for the crisis period (from the fourth quarter of 2006 to the fourth quarter of 2008). *** is significant at 1%, ** is significant at 5%, and * is significant at 10%. t-statistic is reported in the parenthesis.

DepVar:	$CAR(1,30)_{s,t+1}$		$CAR(1,45)_{s,t+1}$		$CAR(1,60)_{s,t+1}$	
$SUERank_{s,t+1}$	0.008***	0.014***	0.010***	0.021***	0.011***	0.026***
	(17.67)	(4.99)	(18.60)	(5.97)	(18.09)	(6.17)
$DummyMember_{s,t}$	-0.005*	0.004	-0.012***	0.003	-0.014***	0.005
	(-1.84)	(1.10)	(-3.74)	(0.59)	(-4.04)	(0.94)
$SUERank_{s,t+1} \times DummyMember_{s,t}$	-0.006***	-0.003***	-0.008***	-0.003***	-0.010***	-0.004***
	(-9.39)	(-2.94)	(-9.77)	(-2.92)	(-10.32)	(-3.20)
Size		-0.031***		-0.048***		-0.062***
		(-19.28)		(-22.21)		(-26.77)
BM		-0.017***		-0.018***		-0.024***
		(-6.56)		(-4.98)		(-6.25)
Momentum		-0.009**		-0.008*		-0.017***
		(-2.40)		(-1.65)		(-3.13)
$SUERank_{s,t+1} \times Size$		-0.001***		-0.001***		-0.001***
		(-3.30)		(-3.74)		(-4.08)
$SUERank_{s,t+1} \times BM$		0.002**		-0.000		-0.000
		(2.52)		(-0.19)		(-0.24)
$SUERank_{s,t+1} \times Momentum$		0.004***		0.006***		0.008***
		(3.60)		(4.62)		(5.57)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year-QTR FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	298,936	298,936	298,936	298,936	298,936	298,936
Adjusted R-Squared	0.0421	0.0527	0.0443	0.0592	0.0454	0.0639

Table A.12: Regress PEAD on non-industry ETF membership

Table A.12 reports the regression result on the following model

$$CAR(1,k)_{s,t+1} = \alpha_s + \alpha_t + \beta_1^S SUERank_{s,t+1} + \beta_2^S DummyNindETFMember_{s,t} + \beta_3^S SUERank_{s,t+1} \times DummyNMember_{s,t} + controls + \epsilon_{s,t}.$$

CAR is cumulative size-adjusted returns for different post-earnings-announcement window. *SUERank* is the quintile ranking of SUE in our sample. *DummyNindETFMember* is a dummy variable, which equals one if the stock is held by at least one non-industry ETF in the current earnings announcement quarter. We control for market capitalization, book-to-market ratio, past 12-month returns and interact them with *SUERank*. In addition, we also control for the year-quarter, firm, and ETF fixed effect. All standard errors are clustered by stock and announcement date.. We run the above regression model on our earnings announcements ranging from 1999 to 2017 except for the crisis period (from the fourth quarter of 2006 to the fourth quarter of 2008). *** is significant at 1%, ** is significant at 5%, and * is significant at 10%. t-statistic is reported in the parenthesis.

DepVar:	$\boldsymbol{CAR(1,30)_{s,t+1}}$	$\boldsymbol{CAR(1,45)_{s,t+1}}$	$\boldsymbol{CAR(1,60)_{s,t+1}}$
$SUERank_{s,t+1}$	0.019***	0.028***	0.034***
	(8.04)	(9.31)	(9.67)
$DummyNindETFMember_{s,t}$	-0.005	-0.011**	-0.015***
	(-1.41)	(-2.57)	(-3.09)
$SUERank_{s,t+1} \times DummyNindETFMember_{s,t}$	-0.001	-0.000	-0.001
	(-0.85)	(-0.21)	(-0.44)
Size	-0.030***	-0.047***	-0.061***
	(-20.48)	(-24.01)	(-28.60)
BM	-0.017***	-0.019***	-0.026***
	(-6.31)	(-5.41)	(-6.51)
Momentum	-0.011***	-0.007	-0.017***
	(-2.89)	(-1.51)	(-3.25)
$SUERank_{s,t+1} \times Size$	-0.001***	-0.002***	-0.002***
	(-6.78)	(-7.71)	(-8.25)
$SUERank_{s,t+1} \times BM$	0.001**	-0.000	-0.001
	(1.99)	(-0.48)	(-0.67)
$SUERank_{s,t+1} \times Momentum$	0.004***	0.006***	0.008***
	(3.99)	(4.50)	(5.65)
Controls	Yes	Yes	Yes
Year-QTR FE	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
Observations	329,386	329,386	329,386
Adjusted R-Squared	0.0549	0.0618	0.0667

Table A.13: Alternative efficiency measures on matched sample

Table A.13 reports regressions of price delay (or variance ratio) on industry ETF membership dummy variable in the match sample. We form the match sample in the same way with Table 11. To compute monthly price delay measures, we regress daily stock returns on contemporaneous and four days of lagged market returns on following model in each stock-month, i.e.,

$$Ret_{s,t}, = \alpha_s + \beta_s^D RM_t + \sum_{n=1}^{4} \delta_{s,n} RM_{t,n} + \epsilon_{s,t}.$$

Ret is the daily return on stock s and *RM* is the daily value-weighted market return. We also run a second regression with restriction that coefficients of lagged market returns are zeros. We define our first price delay measure, *Depay1*, as $1 - R^2(Restricted\ model)/R^2(Unrestricted\ model)$. We define our second price delay measure, *Delay_2*, as $\sum_{n=1}^{4} |\delta_{s,n}| / (|\beta_s^D| + \sum_{n=1}^{4} |\delta_{s,n}|)$, where coefficient estimates are from unrestricted regressions. In addition, we compute quarterly variance ratio, *VR*(*n*,*m*), as the ratio of m-day return variance over n-day return variance multiplied by *n*/*m*. Panel A reports results from monthly regressions of price delay in the next month on industry ETF membership in the current month. *DummyMember*, which equals one if a stock is owned by at least one industry ETF at the beginning of the current month. We control for market capitalization, volume-weighted average price (VWAP), institutional ownership, log of dollar trading volume orthogonalized to size (volume), and one-period lagged dependent variable. we also control for the year-month and industry fixed effect. Standard errors are clustered by stock and year-month. We run the above regression model on our earnings announcements ranging from 1999 to 2017 except for the crisis period (from the fourth quarter of 2006 to the fourth quarter of 2008). Panel B reports results from quarterly regressions of variance ratio in the next quarter on industy ETF membership in the current quarter, and regression specification is similar with Panel A. Standard errors are clustered by stock and year-quarter. *** is significant at 1%, ** is significant at 5%, and * is significant at 10%. t-statistic is reported in the parenthesis.

Panel A: Regress price delay on industry ETF membership in match sample

DepVar:	***Delay_1***	***Delay_2***
$DummyMember_{s,t}$	-0.015***	-0.011***
	(-3.33)	(-4.22)
Size	-0.061***	-0.036***
	(-8.77)	(-8.46)
VWAP	0.008	0.005
	(1.22)	(1.38)
Institutional Ownership	-0.020**	-0.010*
	(-2.26)	(-1.81)
Volume	0.000***	0.000***
	(2.95)	(2.85)
Lag_Dependent	0.040***	0.025***
	(6.69)	(4.42)
Year-Month FE	Yes	Yes
Industry FE	Yes	Yes
Observations	102,327	102,327
Adjusted R-Squared	0.261	0.234

Panel B: Regress variance ratio on industry ETF membership in match sample

DepVar:	*|1-VR(1,5)|*	*|1-VR(1,10)|*	*|1-VR(1,20)|*	*|1-VR(2,5)|*	*|1-VR(2,10)|*	*|1-VR(2,20)|*
$DummyMember_{s,t}$	-0.009***	-0.010***	-0.011**	-0.005**	-0.007**	-0.007*
	(-3.25)	(-2.70)	(-2.64)	(-2.61)	(-2.28)	(-1.79)
Size	-0.025***	-0.020***	-0.020***	-0.015***	-0.014***	-0.018***
	(-6.90)	(-4.71)	(-4.74)	(-5.83)	(-3.56)	(-4.46)
VWAP	0.006**	0.004	-0.004	0.003*	0.002	-0.005
	(2.34)	(1.17)	(-1.13)	(1.78)	(0.68)	(-1.52)
Institutional Ownership	-0.026***	-0.035***	-0.027***	-0.021***	-0.031***	-0.027***
	(-5.52)	(-4.50)	(-3.62)	(-7.10)	(-4.81)	(-3.66)
Volume	-0.000*	-0.000	-0.000*	-0.000*	-0.000	-0.000
	(-1.75)	(-1.39)	(-1.89)	(-1.88)	(-1.58)	(-1.47)
Lag_Dependent	0.045***	0.031***	0.018***	0.019***	0.017**	0.006
	(5.50)	(4.37)	(3.39)	(2.85)	(2.57)	(1.19)
Year-Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	32,061	32,061	32,061	32,061	32,061	32,061
Adjusted R-Squared	0.0381	0.0292	0.0300	0.0331	0.0287	0.0331