Inattentive Price Discovery in ETFs*

Mariia Kosar[1] and Sergei Mikhalishchev[2]

[1,2]CERGE-EI†
[2]Durham University‡

November 4, 2022

Abstract

This paper studies the information choice of exchange-traded funds (ETF) investors, and its impact on the price efficiency of underlying stocks. First, we show that the learning of stock-specific information can occur at the ETF level. Further, our results suggest that ETF investors respond endogenously to changes in the fundamental value of underlying stocks, in line with the rational inattention theory. Second, we provide evidence that ETFs facilitate propagation of idiosyncratic shocks across its constituents.

Keywords: Exchange-Traded Fund, ETF, Price Efficiency, Rational Inattention, Information Acquisition, Comovement
JEL classification codes: G12, G14, D82

---

*This project was supported by Charles University, GAUK project 36119. We are grateful to Jan Hanousek, Filip Matějka, Veronika Selezneva, and Ekaterina Travova for helpful comments and suggestions. All remaining errors are our own.

†CERGE-EI, a joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences, Politickych veznu 7, 111 21 Prague, Czech Republic. Email addresses: mariia.kosar@cerge-ei.cz, sergei.mikhalishchev@cerge-ei.cz.

‡Durham University Business School, Department of Economics and Finance, Mill Hill Lane, Durham, UK, DH1 3LB.

1 Introduction

Exchange-traded funds (ETFs) have gained popularity among investors over the past decades, and have rapidly grown in terms of assets under management and trading volume. These instruments have attracted the attention of both scholars and practitioners due to the important asset pricing implications for their underlying securities. The most well-documented concern about ETFs is their disposition to noise and factor trading that, combined with the continuous arbitrage mechanism, may lead to propagation of noise to the underlying assets (Bhattacharya and O'Hara, 2020). However, there is still a question regarding whether ETFs can facilitate stock-specific price discovery, and if yes what net effect it has for the ETF's underlying bundle.

In this paper we investigate this question. First, we show that the learning of stock-specific fundamental information can occur at the ETF level. Moreover, our results suggest that ETF investors endogenously respond to changes in the fundamental value of underlying stocks, in line with the rational inattention theory[1]. Second, we provide evidence that this pattern of learning affects ETF's underlying bundles, leading to propagation of idiosyncratic shocks across underlying stocks.

We proceed in two steps. Firstly, in order to demonstrate that the information acquisition can occur at the ETF level, we measure the response of ETF intraday prices to earnings surprises. We use earnings surprises as a measure of stock-specific information released at the time of announcement. We focus on capitalization-based ETFs that are traded on U.S. exchanges and have international exposure. We then select only earnings announcements that occur when underlying exchange is closed, and the U.S. exchange is open. By design, this ensures that price discovery, if any, occurs at the ETF level. Moreover, to make ensure that the responses we measure refer to the specified earnings announcements, we select only announcements that were not surrounded by other announcements. Our results suggest that stock specific

[1] A recent review of the rational inattention literature can be found in Maćkowiak et al. (2021).

price discovery can occur at the ETF level. In addition, the earnings response coefficients are statistically significant only for announcements made by firms with large weights in their corresponding ETFs. Furthermore, we differentiate between non-busy days, when there is relatively low news pressure from the U.S. market in terms of macro and stock-specific announcements, and busy information days. On busy days the the response of prices to earnings surprises is significantly smaller.

Secondly, we conduct an empirical analysis of the spillover patterns from ETFs to the stocks in their underlying bundles. Specifically, we compute the abnormal idiosyncratic volatility (AIV) of ETFs and their constituents around earnings announcements (Yang et al., 2020). The AIV measures to which extent the idiosyncratic volatility on announcement days is abnormal compared to the aggregate idiosyncratic volatility over a given period. Then, we estimate the relationship between the AIV of constituent stocks and their corresponding ETFs when the underlying markets re-open. Our results suggest that there is a significantly positive relationship between the AIV of constituent stocks and their corresponding ETFs, which is significant only around earnings releases of stocks with large weights in ETFs. This allows us to conclude that learning at the ETF level affects underlying bundles, leading to abnormal co-movement in volatilities across underlying stocks.

Finally, we show that the ETF AIV risk is priced in a sample of all ETF constituents. The abnormal stock returns are loaded on the ETF AIV, which results in positive and significant regression coefficients of future returns on the ETF AIV over a relatively long time horizon (10 days). The relationship is reversed, which implies that the reaction of returns to the ETF AIV was not fundamental.

Literature. This study contributes to several strands of literature. Firstly, the results in this paper relate to the literature on the impact of financial innovation on the efficiency of financial markets (Basak and Pavlova, 2013; Appel et al., 2016). There is a growing academic literature on the effects of ETFs on the asset pricing of their constituents. Many researchers treat ETFs mostly as venues for noise or factor trading, and thus focus on propagation of non-fundamental and factor shocks from ETFs to underlying markets (Wang and Xu, 2019; Filippou et al., 2019; Ben-David et al., 2018; Shim, 2018; Huang et al., 2021; Israeli et al., 2017; Glosten et al., 2016; Levy and Lieberman, 2019). Two prominent reasons for such concerns are best summarized by Ben-David et al. (2018) and Shim (2018). Ben-David et al. (2018) argue that ETF investors are dominated by noise traders, who propagate non-fundamental shocks to prices of underlying assets, amplifying non-fundamental volatility. Shim (2018) takes a different approach,arguing that ETF markets are populated with informed traders who are, however, factor-informed. He shows that, if factor price discovery occurs in ETFs, rather than stocks, underlying securities tend to misreact to factor information. Both approaches ascribe the key role in shock propagation from ETFs to underlying securities to ETF arbitrage mechanism. However, some studies have reached a conclusion that, due to benefits that such instruments bring to the market (i.e., low cost, high liquidity, and hedging opportunities), ETFs can encourage informed trading and information transfers around fundamental news releases, and thus improve the pricing efficiency of their underlying stocks (Ciura, 2016; Huang et al., 2021; Bhojraj et al., 2020; Ernst, 2021). For example, Bhojraj et al. (2020) focus only on top-weighted stocks and show that ETF mechanic bundle trades help to transfer sector and market-wide information contained in company earnings announcements into the stock prices of its peers, reducing their post-earnings announcement drift and thus contributing to their price efficiency. This is consistent with Savor and Wilson (2016), who show that investors learn both factor and asset-specific components from earnings announcements.

Relative to these studies, we focus on the role of ETFs in transferring an asset's value-specific information to other assets[2]. We use identification strategy, which allows us to study how exactly ETF investors acquire information about their con-

[2]Bhattacharya and O'Hara (2018) theoretically show that ETFs may have a detrimental influence on information propagation from one stock to another, since they can also transfer value-irrelevant firm-specific shocks to their peers, which may lead to market instability and increased synchronicity between stock prices.

stituents, and to evaluate the net effect of price discovery on the ETF level for underlying bundles.

Secondly, this paper is closely related to literature that links asset price responses to investor inattention. While, there are many empirical studies that document this phenomenon (for example, Barber and Odean 2008; Hirshleifer et al. 2009; DellaVigna and Pollet 2009; Fedyk 2021), there is still a lack of empirical literature that studies endogenous investor attention and shows how investors actually behave[3]. Chuprinin et al. (2019) show that firm size is a major determinant of the degree of investor research into a specific stock around fundamental news releases. Li (2022) shows that the efficiency of price reaction to a particular type of risk depends on the value-relevance of that risk. Kacperczyk et al. (2016) demonstrate that mutual fund managers optimally track information about aggregate shocks in recessions and idiosyncratic shocks in booms. Recent studies by Hirshleifer and Sheng (2021) and Huang et al. (2019) investigate how stock investors allocate attention between systematic and idiosyncratic information. We complement this literature by focusing on endogenous investor attention[4]. However, we focus on ETF which, for example, in contrast to mutual funds, have a fixed weighting scheme that allows us to isolate the effect of news releases on changes in the price of ETF, so that we can obtain a clear measure of attention using intraday data[5].

Finally, this project contributes to the strand of literature on the importance of foreign investments into local financial markets (Figlio and Blonigen, 2000; Levy and Lieberman, 2019; Filippou et al., 2019). Specifically, we construct a diverse sample of ETFs that focus on various country and sector indexes. From this diverse sample, we are able to establish the impact of U.S. - traded ETFs on local stocks in

[3]There are numerous theoretical papers that use endogenous inattention to understand comovements or sluggishness of prices, for example Coibion and Gorodnichenko (2012); Mackowiak and Wiederholt (2009); Veldkamp (2006)

[4]See also Ben-David et al. (2021) who document competition for attention in the ETF space by creating specialized ETFs.

[5]Ernst (2021) also studies ETF and presents empirical evidence that simultaneous trades of ETFs with their announcing constituent stocks increase on earnings announcement days, and more so for stocks with high weights in ETFs.

their underlying bundles.

The rest of the paper is organized as follows. In Section 2 we set up a basic theoretical framework of investor's behavior when she faces information constraint. Empirical research design and data are outlined in Sections 3 and 4. Section 5 discusses the results. Finally, Section 6 concludes.

2 Theoretical framework

We model the investor's behavior following the literature on rational inattention, which originated in studies by Sims (2003, 1998). For tractability, we consider a one period two-dimensional tracking problem with quadratic loss[6]. The investor wants to track changes in the value of the ETF: $\Delta V = \sum_i w_i \Delta V_i$, where ΔV_i are changes in the liquidation value of stock $i \in 1, 2$ that enters the ETF with weight $w_i > 0$. However she can process only a finite amount of information. We model the limited ability to process information as a constraint on uncertainty reduction, where uncertainty is measured by entropy (Shannon, 1948; Cover and Thomas, 2012). The problem is formalized as follows.

RI problem. *The investor's problem is to choose the joint distribution of the decision variable ΔV with the exogenous uncertainty ΔV_i, $i \in \{1, 2\}$ so as to maximize:*

$$\max_{\Delta V} \mathbb{E}[-(w_1 \Delta V_1 + w_2 \Delta V_2 - \Delta V)^2],$$

where priors are

$$\forall i \in \{1, 2\} : \quad \Delta V_i \sim N(0, \sigma_i^2).$$

The investor can obtain independent signals about the individual liquidation value of stock i:

$$\forall i \in \{1, 2\} : \quad s_i = \Delta V_i + e_i,$$

[6]We show in Appendix A.2 that results are qualitatively the same for the multi-dimensional tracking problem. Also see Veldkamp (2006) for more general treatment of the problem.

where the noise of signals is normally distributed, $e_i \sim N(0, \sigma^2_{e_i})$. The variance of the signals, $\sigma^2_{e_i}$, is subject to investors choice.

The investor has a capacity constraint in the choice of signal[7]

$$\sum_i \underbrace{\frac{1}{2}\log(\frac{\sigma^2_i}{\sigma^2_{i|s_i}})}_{k_i} \leq k, \tag{1}$$

where $\sigma^2_{i|s_i}$ is a conditional variance of changes in the value of individual stock i, k is the bound on the investor's capacity to process information, and k_i is the investor's attention to value-relevant information of the stock i.

In addition, the investor faces the no-forgetting constraint, i.e., condition that she can not increase prior uncertainty about changes in the stock's value:

$$\sigma^2_{i|s_i} \leq \sigma^2_i. \tag{2}$$

Because priors and noises are normal, $\sigma^2_{i|s_i}$ is a monotone function of $\sigma^2_{e_i}$: $\sigma^2_{i|s_i} = \frac{\sigma^2_{e_i}\sigma^2_i}{\sigma^2_{e_i}+\sigma^2_i}$. In Appendix A.1 we show that the problem of the investor reduces to the choice of $\sigma^2_{i|s_i}$.

The solution to the problem is formalized in the following lemma:

Lemma 1. *The optimal investor's choice of conditional variances of changes in values of individual stocks and attention to value-relevant information of stocks are:*

$$\sigma^2_{i|s_i} = \min\{\sigma^2_i, \frac{w_{\neg i}}{w_i}\sqrt{e^{-2k}\sigma^2_i\sigma^2_{\neg i}}\}$$

$$k_i = \max\{0, \frac{1}{2}\log(\frac{w_i\sqrt{\sigma^2_i}}{w_{\neg i}\sqrt{e^{-2k}\sigma^2_{\neg i}}})\}. \tag{3}$$

[7]This can be motivated by investors having just 168 hours a week. An alternative way to model the behavior is to assume information-processing costs, such that investors may be able to expand their attention whenever needed. Therefore, investors attention to the specific asset will not depend on information that is not directly relevant. Our empirical results (see Section 5.1) could be interpreted as supporting both models. Hence, we remain agnostic on this question, and additional tests are needed to separate these two models. See Azrieli (2021) for a discussion of the difference between model approaches.

Proof. See Appendix A.1. □

Following Lemma (1) and taking derivatives of equation (3) with respect to stock weights, the variance of changes in a stock's value, and an investor's capacity to process information yields the following results:

Corollary 1 (Testable implications). *An investor's attention to a stock's value-relevant information is higher for*

1.1. *stocks with higher relative weights in the ETF;*

1.2. *stocks with higher volatility of changes in the value;*

1.3. *investors with higher information capacity.*

According to Corollary (1.1) the ETF response should be higher for stocks with higher weight in the ETF, controlling for other potential factors. Corollary (1.2) states that, if the volatility of changes is high, which in terms of our empirical exercise means high earnings surprises, then the response of the ETF price will be more efficient. Corollary (1.3) indicates that, if investors have lower information capacity, then the ETF price efficiency with respect to stock information decreases. We test this by comparing the ETF price response in busy days and in days with low numbers of informational announcements.

3 Empirical research design

3.1 ETF-level analysis

Identifying the response to announcements. The most challenging task in our empirical exercise is to identify the response to the earnings announcement shock on ETF level. The first challenge is to isolate the ETF price response to a specific constituent stock earnings announcement. An average ETF contains dozens of stocks which can make concurrent information releases. To attribute the ETF price response to a specific earnings release, it is necessary to ensure that no other constituent in that ETF makes a competing announcement within a chosen time window. To mitigate this problem, we consider only announcements that are not surrounded by competing earnings releases[8] in the same ETF within a [-1 working day, +1 working day] non-announcement window[9].

The second challenge is to attribute the ETF price response to the price discovery on ETF level. Because of the continuous arbitrage process that occurs between ETFs and their underlying bundles, it can be hard to identify where the price discovery occurs, in the ETF or in it's underlying bundle. To mitigate this issue, we consider only ETFs with asynchronous trading hours with their underlying bundles. Those are ETFs that are traded on U.S. exchanges, but have exposure to international markets. For this sample of ETFs, we are able to observe their price responses when the underlying markets are temporarily closed, but the companies on the underlying markets continue to release earnings announcements. Further, to ensure that the ETFs and their underlying markets do not interact during announcement windows, we require at least 6 hours time lapse from an announcement to the next underlying market's opening. This approach allows us to identify ETFs as a source of price discovery, since the arbitrage mechanism is temporarily switched off.

We include fund fixed effect to capture the differences in fund characteristics, mainly the size and liquidity, which can significantly affect the speed and magnitude of fund price response around information releases. Finally, day fixed effect is included to capture the overall market differences common for all stocks and funds, for example, market volatility and information quantity released during a particular day.

Empirical specification. To test if the investor's attention to stock's value-

[8]Although information releases are not limited to earnings announcements, usually other information releases are done together with the earnings releases as a part of quarterly/yearly disclosure.

[9]The choice of a non-announcement window is motivated by sample size considerations, as well as by the empirical literature that usually employs a 3-day window for calculating the announcement price response to earnings announcement on stock level.

relevant information is higher for stocks with higher relative weights in the ETF (Corollary 1.1) and stocks with higher volatility of changes in the value (Corollary 1.2), we measure the response of ETF prices to earnings surprises by computing earnings response coefficients (ERC) over different time horizons for different ETF weight quaniles. ERC present price elasticity with respect to information contained in earnings surprise (Blankespoor et al., 2020), and are obtained from regressing window returns around earnings announcements on earnings surprise. The main empirical model of interest is:

$$ret_{i,j,[\tau',\tau]} = \alpha SUR_{i,j,t} + \beta I_{W_{i,j} \in q} + \gamma SUR_{i,j,t} * I_{W_{i,j} \in q} + \delta_i + \delta_t + Controls_{i,j,t} + \epsilon_{i,j,[\tau',\tau]} \tag{4}$$

where $ret_{i,j,[\tau',\tau]}$ is the cumulative return over announcement window $[\tau', \tau]$; $I_{W_{i,j} \in q}$ is the indicator function that takes the value of 1 if the weight of stock j in the ETF i is in the q^{th} quartile of ETF weights distribution; $SUR_{i,j,t}$ is the earnings surprise; δ_i and δ_t are ETF and day fixed effects. Controls include past period cumulative returns, $ret_{i,j,[\tau'-2,\tau-2]}$, to capture existing time-series dependence of ETF returns; weight of stock j in ETF i, $W_{i,j}$, and the log of market capitalization of stock j on day t, $log(Mkt\ Cap_{j,t})$, to control for the selection criteria for weights assignment within ETF, which are purely market-cap driven.

Savor and Wilson (2016) show that earnings announcements are signals of the future growth prospects of the firm, and use them as firm-level information events. We follow Hirshleifer et al. (2009) and define the earnings surprise of stock specific announcement j in the ETF i on day t as:

$$SUR_{i,j,t} = \frac{Earnings_{i,j,t} - 1/K \sum_{k=1}^{K} Earnings_{i,k,t}}{P_{i,j,t}}, \tag{5}$$

where $P_{i,j,t}$ is a closing price of stock j in the ETF i on day t, and a mean forecast of earnings of all K analysts for announcement j in the last quarter prior to announcement j is $1/K \sum_{k=1}^{K} Earnings_{i,k,t}$.

We calculate the return $ret_{i,j,[\tau',\tau]}$ over announcement window $[\tau', \tau]$ as:

$$ret_{i,j,[\tau',\tau]} = \sum_{t=\tau'}^{\tau} ret_{i,j,t} = \sum_{t=\tau'}^{\tau} \{log(P_{i,j,t}) - log(P_{i,j,t-1})\}, \tag{6}$$

where $P_{i,j,t}$ is price of the ETF i t minutes past (before) the announcement of stock j; t spans from $\tau' = -4$ to $\tau = +4$ hours with 30 minute intervals.

To ensure that we correctly measure the weight percentile of each announcing stock j in the ETF i, we compute the respective weight percentiles in the full sample of each ETF i constituents on day t.

To investigate whether the investor's attention to stock's value-relevant information is higher for investors with higher information capacity (Corollary 1.3), we study the ERCs on the busy vs. normal days on the U.S. stock market. The empirical model of interest is the following:

$$\begin{gathered} ret_{i,j,[\tau',\tau]} = \alpha SUR_{i,j,t} + +\beta SUR_{i,j,t} * I_{W_{i,j} \in q} + \gamma SUR_{i,j,t} * BUSY_t + \\ \theta SUR_{i,j,t} * I_{W_{i,j} \in q} * BUSY_t + Controls_{i,j,t} + \delta_i + \delta_t + \epsilon_{i,j,[\tau',\tau]}. \end{gathered} \tag{7}$$

where busy day indicator variable $BUSY_t$ is defined as:

$$BUSY_t = 1 \; if \; News \; Score_t > Q_{0.5} \; or \; N > Q_{0.5}. \tag{8}$$

In the above formula, $News \; Score_t$ is the macroeconomic news score of each trading day, and is computed following the methodology of Xu et al. (2018):

$$News \; Score_t = \frac{1}{N} \sum_{1}^{N} Score_{j,t},$$

where $Score_{j,t} = \frac{ESS_{j,t} - 50}{50}$ is the normalized Event Sentiment Score ($ESS_{j,t}$) for event j on day t on U.S. market. The Event Sentiment Score indicates the extent to which an event can influence a market price. N is the total number of news events on U.S. market on day t.

3.2 Stock-level analysis

Empirical specification. In this section, we introduce an empirical model to test whether learning patterns at the ETF level spill over to their underlying portfolios through instant arbitrage between ETFs and their constituents after underlying markets re-open following announcements. We adopt the approach of Yang et al. (2020), who introduce the abnormal idiosyncratic volatility (AIV) as a measure of information risk associated with earnings announcements. The AIV measures the extent to which the idiosyncratic volatility on announcement days is abnormal compared to the aggregate idiosyncratic volatility over a given period. We consider quarterly earnings announcements and, hence, use a quarter period. To measure the AIV of constituent stocks, for each unique ETF i within our sample of fund-announcement data, we collect data on all constituent stocks during our sample period (2016-2017). For each of these stocks, we use data on Fama-French factors, and estimate the idiosyncratic returns with a 3 factor Fama-French model using daily data:

$$ret_{j,t} = \alpha_j + \beta_j^{MKT} MKT_t + \beta_j^{SMB} SMB_t + \beta_j^{HML} HML_t + \epsilon_{j,t}, \tag{9}$$

where $ret_{j,t}$ are close-to-close returns stock j from day $t-1$ to t; MKT is the value-weighted market portfolio excess return over the risk-free rate; SMB is the size factor; and HML is the value factor; and $\epsilon_{j,t}$ is the abnormal idiosyncratic return.

Next, for each stock j that entered fund i during the announcement day t we compute the idiosyncratic volatility of a stock within a quarter for the announcement days (IV^{AD}), which are the trading session before an announcement that occurred during off-exchange hours, and the next two trading sessions after the announcement; and for non-announcement days (IV^{NAD}) as the log of the standard deviations of the residual from equation (9) during these days, assuming that there are 63 trading days in a quarter. More specifically, we define:

$$IV^{AD}_{j,i,t} = ln\sqrt{\frac{63 * \sum_{t \in AD} \epsilon^2_{j,i,t}}{(n_{AD} - 1)}},$$

$$IV^{NAD}_{j,i,t} = ln\sqrt{\frac{63 * \sum_{t \in NAD} \epsilon^2_{j,i,t}}{(n_{NAD} - 1)}},$$

where n_{AD} and n_{NAD} are the number of days in the pre- and non-announcement periods, respectively. We compute the AIV around announcement day t as the difference in log idiosyncratic volatility:

$$AIV_{j,i,t} = IV^{AD}_{j,i,t} - IV^{NAD}_{j,i,t}.$$

Similarly, we estimate the equation (9) for returns of ETF i from day $t-1$ to t:

$$ret_{i,t} = \alpha_i + \beta_i^{MKT} MKT_t + \beta_i^{SMB} SMB_t + \beta_i^{HML} HML_t + \epsilon_{i,t},$$

where $ret_{i,t}$ are close-to-close returns of fund i from day $t-1$ to t; and $\epsilon_{i,t}$ is the abnormal idiosyncratic return. We compute the idiosyncratic volatility of ETF i for announcement days (IVF^{AD}), which are the trading day on the U.S. market before an announcement of stock j occurred, the trading day on the U.S. market on which the announcement of stock j occurred, and the next day after the announcement; and for non-announcement days (IVF^{NAD}) in a given quarter as:

$$IVF^{AD}_{j,i,t} = ln\sqrt{63 * \frac{\sum_{t \in AD} \epsilon^2_{i,t}}{(n_{AD} - 1)}},$$

$$IVF^{NAD}_{j,i,t} = ln\sqrt{63 * \frac{\sum_{t \in NAD} \epsilon^2_{i,t}}{(n_{NAD} - 1)}}.$$

Then, the AIV of ETF i around announcement day t is:

$$AIV_{i,t} = IVF^{AD}_{j,i,t} - IVF^{NAD}_{j,i,t}.$$

We measure the relationship between the AIV of ETFs around the announcement and the AIV of its constituent non-announcing stocks during the next open trading session after the announcement. The empirical model is:

$$AIV_{j,i,t} = \alpha_i + \alpha_j + \alpha_t + \sum_q \gamma_q AIV_{i,t} I_{W_{i,k} \in q} + Controls_{j,t} + \eta_{j,i,t}, \tag{10}$$

where $AIV_{j,i,t}$ is the abnormal idiosyncratic volatility of stock j in ETF i on announcement day t; $AIV_{i,t}$ is the abnormal idiosyncratic volatility of ETF i on announcement day t on the U.S. market; $I_{W_{i,k}}$ is the indicator for announcing stock k in ETF i on announcement day t being in the q^{th} quartile of the ETF weights distribution. Following Ben-David et al. (2018) and Yang et al. (2020), $Controls_{j,t}$ include the inverse of price of stock j on day t, $\frac{1}{P_{j,t}}$, the log of market capitalization of stock j on day t, $log(Mkt\ Cap_{j,t})$, the log of Amihud illiquidity measure of stock j on day t, $log(Amihud_{j,t})$, and the lagged returns ($ret_{j,[-1]}$, $ret_{j,[-3,-2]}$, $ret_{j,[-6,-4]}$).

Mechanisms. Bhattacharya and O'Hara (2018) theoretically show that shocks from ETFs are transmitted at a higher degree to the stocks with higher weights in ETFs. To test this empirically, we add the weight of stock j in the ETF i on day t, $W_{i,j,t}$ to our empirical specification, and estimate the following model:

$$\begin{aligned} AIV_{j,i,t} = \alpha_i + \alpha_j + \alpha_t + \sum_q \gamma_q AIV_{i,t} I_{W_{i,k} \in q} + \sum_q \beta_q log(W_{i,j,t}) I_{W_{i,k} \in q} \\ + \sum_q \beta_q AIV_{i,t} log(W_{i,j,t}) I_{W_{i,k} \in q} + Controls_{j,t} + \eta_{j,i,t}. \end{aligned} \tag{11}$$

Following Ben-David et al. (2018) and Shim (2018), we also test whether the arbitrage trades that occurs between ETFs and their underlying bundles can explain the correlations between stocks and ETFs. The model of interest is as follows:

$$\begin{aligned} AIV_{j,i,t} = \alpha_i + \alpha_j + \alpha_t + \sum_q \gamma_q AIV_{i,t} I_{W_{i,k} \in q} + \sum_q \beta_q \Delta_{i,j,t} I_{W_{i,k} \in q} \\ + \sum_q \beta_q AIV_{i,t} \Delta_{i,j,t} I_{W_{i,k} \in q} + Controls_{j,t} + \eta_{j,i,t}. \end{aligned} \tag{12}$$

We compute the intensity of arbitrage, $\Delta_{i,j,t}$, as the normalized change in the number of the total shares of stock j held by each ETF i on day t:

$$\Delta_{i,j,t} = |\frac{Shares_{j,i,t} - Shares_{j,i,t-1}}{Shares_{j,t}}|,$$

where $Shares_{j,i,t}$ is the number of shares of stock j held by ETF i on day t; $Shares_{j,t}$ is the total number of shares of stock j on day t.

Further evidence. Finally, we test whether the AIV of ETF is priced. We follow Eugene and French (1992) and estimate the following regression:

$$aret_{j,[t,t+m]} = a + b * AIV_{j,i,t} + \sum_q \gamma_q AIV_{j,i,t} I_{W_{i,k} \in q} + Controls_{j,t} + \epsilon_{j,t}, \tag{13}$$

where $aret_{j,[t,t+m]}$ is stock j's cumulative abnormal return, which is the sum of abnormal daily returns, $\epsilon_{j,t}$, from the announcement on day t to day t+m; $Controls_{j,t}$ are the same as in previous regressions.

4 Data

4.1 ETF-level data

Data on daily ETF constituents and their weights in each ETF comes from the ETFDB database. We start with an initial ETF sample that includes all U.S. -traded capitalization-based ETFs with international exposure that active during 2016-2017. We obtain the respective ETF tickers from etf.com. We exclude all sector ETFs from our initial sample, and keep only ETFs with country and regional exposure. Within each ETF, we split all constituent stocks into percentiles by their corresponding weight in the ETF.

To construct a measure of surprise earnings, we collect data on quarterly earnings announcements from I/B/E/S for each ETF constituent. Specifically, we retrieve the following variables from I/B/E/S: the date and time of each announcement, official tickers of the announcing stocks, announced earnings per share (EPS) and the analyst forecasts of EPS for each announcement. The I/B/E/S and ETFDB are matched based on constituent CUSIP.

We obtain daily prices of each announcing ETF constituent from Compustat Daily International. We use this data to compute the earnings surprises. Compustat and I/B/E/S data are matched based on a 6-digit CUSIP obtained from the 8-digit CUSIP in I/B/E/S and from SEDOL in Compustat.

Data on the off-exchange hours of the underlying ETF markets and the opening hours of the U.S. exchanges comes from tradinghours.com. Moreover, we require that there is [-1 day, +1 day] non-announcement window around each announcement. We also ensure that there is at least 6 hours after the announcement prior to the underlying market opening. This procedure leaves us with 842 unique fund-announcement observations.

Data on high-frequency intra-day ETF prices comes from the Trades and Quotes database. We use intra-day trades data to find all trades made during each announcement day. TAQ trades include information on the date and exact time of a trade (up to a millisecond), and data on the prices and sizes of trading orders. We sample the trades data at 5 second frequency. We keep the last price in each 5 second interval, and sum up all trades made during the respective interval to compute the trading volume. Finally, we use price adjustment factors from the Compustat Quarterly database to account for stock splits.

We use the full Dow Jones Edition of the RavenPack News Analytics database to compute the news score of each trading day. Out of all macroeconomic news related to topics of business and economics, we select those with the highest relevance (Event Novelty Score = 100).

Summary statistics appears in Table 5 in Appendix B.

4.2 Stock-level data

We use data on 842 unique fund-announcement observations from in Section 3.2. For each announcement, we identify all ETFs that hold the announcing stock, and all constituents of such ETFs at the time of announcement. Next, we use the Compustat International daily data on prices and shares outstanding of all identified ETF constituents during period of 2016-2017.

We use the ETFDB data on weights and number of shares held by ETFs to compute the intensity of arbitrage, $\Delta_{i,j,t}$, and the weight of each stock in each specific ETF.

Summary statistics appears in Table 6 in Appendix B.

5 Results

5.1 ETF-level results

Table 1 shows the estimation results of the empirical model in Equation 4. The response in returns occurs only around the information release, which is in accordance with the literature on stock market information processing (Kim and Verrecchia, 1997; Bamber et al., 2011; Back et al., 2018; Yang et al., 2020). As estimates suggest the response of ETF returns to announcements is strongest for stocks in the top percentiles of ETF weight distribution. Specifically, the coefficients of the interaction terms $SUR * 1_{Weight>Q_{0.75}}$ and $SUR * 1_{Weight\in[Q_{0.5},Q_{0.75}]}$ become significant and positive in 4 hours window around the announcement[10]. At the same time, there is no significant effect of the announcement on stocks with weight in other quartiles of the distribution. This means that the higher the weight of the stock the earlier and more efficiently traders would react to information about it.

These results provide a strong evidence in support of Corollaries 1.1 and 1.2: investors rationally adjust their attention in response to earnings announcements. These findings cannot be explained by liquidity and transaction costs, because we control for time and fund specific factors. Moreover, they are not consistent with the salience explanation that investors' attention is drawn to those earnings surprises which are most different relative to the average (Bordalo et al., 2013).

We evaluate the model in Equation 7 to test Corollary 1.3: traders with less cognitive capacity acquire less information. The results are presented in Table 2. We find the similar pattern in return responses: only stocks with higher weights in ETFs respond around the information release. Notably, the results suggest that there is a significant difference in responses to announcements occurred on busy and non-busy days. On busy days traders react less to announcements. Hence, the results are consistent with Corollary 1.3, as well as with the distraction effect theory (Hirshleifer et al., 2009): the arrival of extraneous news causes prices to react sluggishly to relevant news about a firm. However, we cannot distinguish between purely rational and behavioral explanations and, therefore, further research is needed.

[10]The coefficient 0.047 means that one unit increase in earnings surprise leads on average to $\approx 5\%$ increase in log ETF return ($e^{0.047} = 1.048$).

Table 1: Earnings response coefficients of ETFs around earnings announcements

	Dependent variable:				
	$ret_{[-4;-2)}$	$ret_{[-2,0]}$	$ret_{(0,+2]}$	$ret_{(+2,+4]}$	$ret_{(+4,+6]}$
SUR	0.004	-0.005	-0.005	0.001	0.006
	(0.006)	(0.006)	(0.005)	(0.014)	(0.012)
$SUR * 1_{Weight \in [Q_{0.25}, Q_{0.5}]}$	-0.007	0.012	-0.004	0.015	-0.008
	(0.017)	(0.016)	(0.010)	(0.015)	(0.015)
$SUR * 1_{Weight \in [Q_{0.5}, Q_{0.75}]}$	0.018	0.040**	0.002	0.014	-0.018
	(0.036)	(0.019)	(0.008)	(0.017)	(0.020)
$SUR * 1_{Weight > Q_{0.75}}$	-0.072	0.021	0.047**	-0.060	-0.055
	(0.050)	(0.038)	(0.019)	(0.074)	(0.080)
Controls	Yes	Yes	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes	Yes	Yes
Note:			*p<0.1; **p<0.05; ***p<0.01		

Note: This table presents estimates of regression of ETF returns over specified announcement windows ([-4; -2),[-2, +2], (+2, +4] hours around the announcement) on a measure of earnings surprise of a stock within corresponding ETF, SUR. $I_{W/\in q}$ is the indicator function that takes the value of 1 if the weight of stock in the ETF is in the q^{th} quartile of ETF weights distribution. $Controls$ include the log of market capitalization of non-announcing stock, $log(Mkt\ Cap)$, the weight of announcing stock in the ETF, W, and the last window lag return, $ret_{i,j,[\tau'-2,\tau-2]}$. We use fund and day fixed effects. Standard errors are adjusted for small sample (Arellano et al., 1987) and reported in parentheses. The description of variables is in Section 3.1. The sample period is 2016-2017.

Table 2: Earnings response coefficients of ETFs around earnings announcements - normal vs. busy days

	Dependent variable:				
	$ret_{[-4;-2)}$	$ret_{[-2,0]}$	$ret_{(0,+2]}$	$ret_{(+2,+4]}$	$ret_{(+4,+6]}$
SUR	-0.096	0.016	-0.007	0.114**	-0.024
	(0.064)	(0.044)	(0.024)	(0.051)	(0.052)
$SUR * 1_{Weight \in [Q_{0.25}, Q_{0.5}]}$	0.054	-0.014	0.024	-0.033	0.077
	(0.049)	(0.039)	(0.017)	(0.048)	(0.053)
$SUR * 1_{Weight \in [Q_{0.5}, Q_{0.75}]}$	-0.013	0.057	-0.008	-0.014	0.082
	(0.179)	(0.056)	(0.033)	(0.066)	(0.052)
$SUR * 1_{Weight > Q_{0.75}}$	-0.306	0.390	0.135**	-0.242	0.066
	(0.196)	(0.252)	(0.054)	(0.214)	(0.123)
$SUR * BUSY$	0.101	-0.022	0.001	-0.115**	0.031
	(0.065)	(0.046)	(0.025)	(0.054)	(0.055)
$SUR * 1_{Weight \in [Q_{0.25}, Q_{0.5}]} * BUSY$	-0.060	0.026	-0.032	0.045	-0.091
	(0.050)	(0.042)	(0.020)	(0.051)	(0.057)
$SUR * 1_{Weight \in [Q_{0.5}, Q_{0.75}]} * BUSY$	0.035	-0.018	0.010	0.026	-0.105*
	(0.185)	(0.058)	(0.035)	(0.070)	(0.059)
$SUR * 1_{Weight > Q_{0.75}} * BUSY$	0.272	-0.412	-0.106*	0.171	-0.159
	(0.208)	(0.257)	(0.061)	(0.224)	(0.197)
Controls	Yes	Yes	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes	Yes	Yes
Note:			*p<0.1; **p<0.05; ***p<0.01		

Note: This table presents estimates of regression of ETF returns over specified announcement windows ([-4; -2),[-2, +2], (+2, +4] hours around the announcement) on a measure of earnings surprise of a stock within corresponding ETF, SUR. Other variables are: the indicator function that takes the value of 1 if the weight of stock in the ETF is in the q^{th} quartile of ETF weights distribution, $I_{W/\in q}$; dummy variable that takes the value 1 when the average news score or the number of relevant events on the U.S. market that day is larger than median, $BUSY$. $Controls$ include the log of market capitalization of non-announcing stock, $log(Mkt\ Cap)$, the weight of announcing stock in the ETF, W, and the last window lag return, $ret_{[\tau'-2,\tau-2]}$. We use fund and day fixed effects. Standard errors are adjusted for small sample (Arellano et al., 1987) and reported in parentheses. The description of variables is in Section 3.1. The sample period is 2016-2017.

5.2 Stock-level results

Table 3 shows the results of the estimation of the empirical models (10)-(12). They suggest that the AIV of non-announcing constituent stocks increases with the AIV of their corresponding ETFs. Moreover, the effect is significant only around announcements for stocks that are above the 75^{th} percentile of ETF weight distribution, which is consistent with the results above. This result holds across all three models and suggests that ETFs could be a source of increased idiosyncratic stock-level volatility that is transferred to the underlying stocks.

We do not find any evidence that the weights of constituents stocks influence the relationship between the AIV of stocks and ETFs (column 2), nor do arbitrage trades of Authorized Participants (column 3). It contradicts the findings of Ben-David et al. (2017), who show that ETF-level shocks are translated to underlying stocks at larger magnitudes if the stock has greater weight in the ETF. However, Bhattacharya and O'Hara (2018) outline a possible no-arbitrage mechanism - direct learning from ETFs prices by stock investors that potentially could be observed in our setting too.

Additionally, the results of the estimation of the model (13) in Table 7 show that the AIV of ETFs is priced in the higher quartiles of ETF weights distribution. That is, the abnormal stock returns are loaded on the ETF's AIV, which is visible from the positive and significant coefficients on the AIV of stocks with weights primarily in fourth quartile. The relationship is then reversed after 10 days, implying that abnormal returns overreact to the ETF AIV, and the overreaction is subsequently corrected.

Table 3: The effect of the AIV of ETFs on the AIV of non-announcing stocks

	Dependent variable: $AIV\ of\ non-announcing\ stock$		
	(1)	(2)	(3)
AIV	0.002	-0.0004	-0.003
	(0.004)	(0.004)	(0.004)
$AIV*1_{W\in[Q_{0.25},Q_{0.5}]}$	-0.003	0.0002	0.001
	(0.004)	(0.004)	(0.004)
$AIV*1_{W\in[Q_{0.5},Q_{0.75}]}$	0.005	0.006^{*}	0.004
	(0.004)	(0.004)	(0.004)
$AIV*1_{W>Q_{0.75}}$	0.020^{***}	0.022^{***}	0.021^{***}
	(0.004)	(0.004)	(0.005)
$AIV*W$		0.007^{**}	
		(0.003)	
$AIV*W*1_{W\in[Q_{0.25},Q_{0.5}]}$		-0.015^{**}	
		(0.006)	
$AIV*W*1_{W\in[Q_{0.5},Q_{0.75}]}$		-0.005	
		(0.004)	
$AIV*W*1_{W>Q_{0.75}}$		-0.008^{**}	
		(0.004)	
$AIV*\Delta$			0.588
			(0.778)
$AIV*\Delta*1_{W\in[Q_{0.25},Q_{0.5}]}$			-2.324
			(1.594)
$AIV*\Delta*1_{W\in[Q_{0.5},Q_{0.75}]}$			-2.053^{***}
			(0.792)
$AIV*\Delta*1_{W>Q_{0.75}}$			-0.888
			(0.839)
Controls	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes
Observations	174,789	259,235	434,024
R^2	0.220	0.236	0.204

*p<0.1; **p<0.05; ***p<0.01

Note: This table presents estimates of regression of the non-announcing stock abnormal idiosyncratic volatility on the abnormal idiosyncratic volatility of ETF during the next open trading session after the announcement (AIV). Other variables are: the indicator for announcing stock in ETF on announcement day being in the q^{th} quartile of ETF weights distribution, I_W; the weight of non-announcing stock in ETF, W; intensity of arbitrage, Δ. $Controls$ include the inverse of price of non-announcing stock, $\frac{1}{P}$, the log of market capitalization of non-announcing stock, $log(Mkt\ Cap)$, the log of Amihud illiquidity measure of non-announcing stock, $log(Amihud)$, and the lagged returns of non-announcing stock ($ret_{[-1]}$, $ret_{[-3,-2]}$, $ret_{[-6,-4]}$). We use fund, stock, and day fixed effects. Standard errors are clustered at stock level and reported in parentheses. The description of variables is in Section 3.2. The sample period is 2016-2017.

Table 4: The effect of the AIV of non-announcing stock on abnormal stock returns

	Dependent variable: Fama-French Adjusted Cumulative Returns				
	$aret_{[t,t+1]}$	$aret_{[t,t+5]}$	$aret_{[t,t+10]}$	$aret_{[t,t+20]}$	$aret_{[t,t+30]}$
AIV	-0.0003**	-0.0003	-0.001***	-0.003***	-0.006***
	(0.0002)	(0.0002)	(0.0003)	(0.001)	(0.001)
$AIV*1_{W\in[Q_{0.25},Q_{0.5}]}$	0.0002	0.0002	0.001***	0.012***	0.015***
	(0.0002)	(0.0003)	(0.0004)	(0.001)	(0.002)
$AIV*1_{W\in[Q_{0.5},Q_{0.75}]}$	0.0003**	0.0001	0.001	-0.002*	0.001
	(0.0002)	(0.0003)	(0.0004)	(0.001)	(0.001)
$AIV*1_{W>Q_{0.75}}$	0.001***	0.001***	0.003***	0.0003	-0.002
	(0.0002)	(0.0003)	(0.0004)	(0.001)	(0.002)
Controls	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	466,440	466,421	445,708	423,597	412,269
R^2	0.108	0.110	0.194	0.115	0.104

*p<0.1; **p<0.05; ***p<0.01

Note: This table presents estimates of regression of the non-announcing stock cumulative abnormal return for a given period on the abnormal idiosyncratic volatility of non-announcing stock within a given ETF (AIV). I_W is the indicator for announcing stock in ETF on announcement day being in the q^{th} quartile of ETF weights distribution. *Controls* include the inverse of price of non-announcing stock, $\frac{1}{P}$, the log of market capitalization of non-announcing stock, $log(Mkt\ Cap)$, the log of Amihud illiquidity measure of non-announcing stock, $log(Amihud)$, and the lagged returns of non-announcing stock ($ret_{[-1]}$, $ret_{[-3,-2]}$, $ret_{[-6,-4]}$). We use fund, stock, and day fixed effects. Standard errors are clustered at stock level and reported in parentheses. The description of variables is in Section 3.2. The sample period is 2016-2017.

6 Conclusion

In this paper, we show that ETFs can be venues for stock-specific price discovery, and that their learning patterns of stock-specific information are consistent with rational inattention theory. Further, we show that these learning patterns are transferred to underlying bundles of ETFs, leading to increased price co-movements in constituent stocks around information releases. Therefore, stock specific shocks in the ETF can affect underlying market prices, even when such information is irrelevant for a particular underlying asset and, hence, it could lead to greater volatility overall. These results suggest that even rational behavior of constrained individuals combined with the design of the new financial instruments could be a potential weakness for the system, and should be taken into account when thinking about future regulations.

We highlight several directions for future investigation. First, while this paper provides empirical evidence suggesting that ETF prices reflect stock-specific information, it is not entirely clear why investors would trade ETFs instead of stocks around stock-specific news releases. One possible explanation is that ETFs are simply more liquid (Ernst, 2021). Future work could analyze the liquidity of ETFs around earnings announcements of their constituents to shed light on this question.

Second, we do not find evidence that arbitrage explains information transfer from ETFs to underlying bundles. Therefore, it could be interesting to explore alternative mechanisms of information transfer, which could explain the learning process fully.

Finally, we find evidence of rational endogenous information acquisition. At the same time, the result, that there is a higher response of window returns to the earnings surprise on non-busy days, is consistent with behavioral inattention theories (Hirshleifer et al., 2009; Bordalo et al., 2013). Moreover, there is a question of whether investors face information costs or constraints (Azrieli, 2021). Exploring and distinguishing different forces behind these results could be a fruitful direction for future research.

References

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

Arellano, M. et al. (1987). Computing robust standard errors for within-groups estimators. *Oxford bulletin of Economics and Statistics 49*(4), 431–434.

Azrieli, Y. (2021). Constrained versus unconstrained rational inattention. *Games 12*(1), 3.

Back, K., K. Crotty, and T. Li (2018). Identifying information asymmetry in securities markets. *The Review of Financial Studies 31*(6), 2277–2325.

Bamber, L. S., O. E. Barron, and D. E. Stevens (2011). Trading volume around earnings announcements and other financial reports: Theory, research design, empirical evidence, and directions for future research. *Contemporary Accounting Research 28*(2), 431–471.

Barber, B. M. and T. Odean (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies 21*(2), 785–818.

Basak, S. and A. Pavlova (2013). Asset prices and institutional investors. *American Economic Review 103*(5), 1728–58.

Ben-David, I., F. Franzoni, and R. Moussawi (2017). Exchange-traded funds. *Annual Review of Financial Economics 9*, 169–189.

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

Ben-David, I., F. A. Franzoni, B. Kim, and R. Moussawi (2021). Competition for Attention in the ETF Space.

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

Bhattacharya, A. and M. O'Hara (2020). *ETFs and systemic risks*. CFA Institute Research Foundation.

Bhojraj, S., P. Mohanram, and S. Zhang (2020). ETFs and information transfer across firms. *Journal of Accounting and Economics 70*(2-3), 101336.

Blankespoor, E., E. deHaan, and I. Marinovic (2020). Disclosure processing costs, investors' information choice, and equity market outcomes: A review. *Journal of Accounting and Economics 70*(2-3), 101344.

Bordalo, P., N. Gennaioli, and A. Shleifer (2013). Salience and asset prices. *American Economic Review 103*(3), 623–28.

Chuprinin, O., A. Gorbenko, and C. M. Kang (2019). Rationally neglected stocks.

Ciura, B. (2016). How Market Catalysts Affect ETFs. *ETFdb.com*.

Coibion, O. and Y. Gorodnichenko (2012). What can survey forecasts tell us about information rigidities? *Journal of Political Economy 120*(1), 116–159.

Cover, T. M. and J. A. Thomas (2012). *Elements of information theory*. John Wiley & Sons.

DellaVigna, S. and J. M. Pollet (2009). Investor inattention and friday earnings announcements. *The Journal of Finance 64*(2), 709–749.

Ernst, T. (2021). Stock-Specific Price Discovery From ETFs.

Eugene, F. and K. French (1992). The cross-section of expected stock returns. *Journal of Finance 47*(2), 427–465.

Fedyk, A. (2021). Front page news: The effect of news positioning on financial markets.

Figlio, D. N. and B. A. Blonigen (2000). The effects of foreign direct investment on local communities. *Journal of Urban Economics 48*(2), 338–363.

Filippou, I., A. Gozluklu, and H. Rozental (2019). ETF Arbitrage and International Diversification.

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

Hirshleifer, D., S. S. Lim, and S. H. Teoh (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *The Journal of Finance 64*(5), 2289–2325.

Hirshleifer, D. and J. Sheng (2021). Macro news and micro news: Complements or substitutes?

Huang, S., Y. Huang, and T.-C. Lin (2019). Attention allocation and return co-movement: Evidence from repeated natural experiments. *Journal of Financial Economics 132*(2), 369–383.

Huang, S., M. O'Hara, and Z. Zhong (2021). Innovation and informed trading: Evidence from industry ETFs. *The Review of Financial Studies 34*(3), 1280–1316.

Israeli, D., C. M. Lee, and S. A. Sridharan (2017). Is there a dark side to exchange traded funds? an information perspective. *Review of Accounting Studies 22*(3), 1048–1083.

Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp (2016). A rational theory of mutual funds' attention allocation. *Econometrica 84*(2), 571–626.

Kim, O. and R. E. Verrecchia (1997). Pre-announcement and event-period private information. *Journal of accounting and economics 24*(3), 395–419.

Levy, A. and O. Lieberman (2019). Overreaction of Country ETFs to US Market Returns: Intraday vs. Daily Horizons and the Role of Synchronized Trading.

Li, J. (2022). Endogenous inattention and risk-specific price underreaction in corporate bonds. *Journal of Financial Economics 145*(2), 595–615.

Maćkowiak, B., F. Matějka, and M. Wiederholt (2021). Survey: Rational inattention, a disciplined behavioral model.

Mackowiak, B. and M. Wiederholt (2009). Optimal sticky prices under rational inattention. *American Economic Review 99*(3), 769–803.

Savor, P. and M. Wilson (2016). Earnings announcements and systematic risk. *The Journal of Finance 71*(1), 83–138.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal 27*(3), 379–423.

Shim, J. J. (2018). Arbitrage comovement.

Sims, C. A. (1998). Stickiness. In *Carnegie-Rochester Conference Series on Public Policy*, Volume 49, pp. 317–356. Elsevier.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics 50*(3), 665–690.

Veldkamp, L. L. (2006). Information markets and the comovement of asset prices. *The Review of Economic Studies 73*(3), 823–845.

Wang, H. and L. Xu (2019). Do exchange-traded fund flows increase the volatility of the underlying index? evidence from the emerging market in china. *Accounting & Finance 58*(5), 1525–1548.

Xu, L., X. Yin, and J. Zhao (2018). Are authorized participants of exchange-traded funds informed traders?

Yang, Y. C., B. Zhang, and C. Zhang (2020). Is information risk priced? evidence from abnormal idiosyncratic volatility. *Journal of Financial Economics 135*(2), 528–554.

A Proofs

A.1 Proof of Lemma 1

We start by solving the maximization problem for given exogenous signals s_1 and s_2:

$$\max_{\Delta V} \mathbb{E}[-(w_1 \Delta V_1 + w_2 \Delta V_2 - \Delta V)^2 | s_1, s_2]. \tag{14}$$

The first order condition is:

$$\Delta V^* = \mathbb{E}[w_1 \Delta V_1 + w_2 \Delta V_2 | s_1, s_2].$$

Then we plug optimal ΔV^* into equation (14) and obtain:

$$\mathbb{E}[(w_1 \Delta V_1 + w_2 \Delta V_2 - \mathbb{E}[w_1 \Delta V_1 + w_2 \Delta V_2 | s_1, s_2])^2 | s_1, s_2]$$

$$= -w_1^2 Var[\Delta V_1 | s_1] - w_2^2 Var[\Delta V_2 | s_2]$$

$$= -w_1^2 \sigma_{1|s_1}^2 - w_2^2 \sigma_{1|s_2}^2.$$

Therefore, now we can reformulate the maximization problem in terms of conditional variances of changes in the values of individual stocks:

$$\max_{\sigma_{1|s_1}^2, \sigma_{1|s_2}^2} -w_1^2 \sigma_{1|s_1}^2 - w_2^2 \sigma_{1|s_2}^2, \tag{15}$$

subject to (1) and (2).

From the constraint (1) we obtain $\sigma_{1|s_1}^2 = e^{-2k} \frac{\sigma_1^2 \sigma_2^2}{\sigma_{2|s_2}^2}$ and substitute it to the maximization function (15):

$$\max_{\sigma_{2|s_2}^2} -w_1^2 \sigma_{2|s_2}^2 - w_2^2 e^{-2k} \frac{\sigma_1^2 \sigma_2^2}{\sigma_{2|s_2}^2}.$$

The first order conditions yields:

$$\sigma^2_{1|s_1} = \frac{w_2}{w_1}\sqrt{e^{-2k}\sigma_1^2\sigma_2^2}$$

$$\sigma^2_{2|s_2} = \frac{w_1}{w_2}\sqrt{e^{-2k}\sigma_1^2\sigma_2^2}.$$

Then we apply the non-forgetting constraint (2) and obtain Lemma 1.

A.2 The multi-dimensional rational inattention problem

Above, we considere the two-dimensional problem. The only difference now is that an ETF consists of $N \in \mathbb{R}$ independent stocks with weights w_i, $i \in 1, ..., N$. Following the same steps as in Appendix A.1, it is easy to show that the solution to this problem is:

$$\forall i \in 1, ..., N: \quad \sigma^2_{i|s_i} = \sqrt[N]{\frac{\prod_{j=1}^N w_2^2}{w_i^4} e^{-2k} \prod_{j=1}^N \sigma_i^2}.$$

Therefore, the comparative statics results are similar to the two-dimensional problem, and hence the latter could be considered without loss of generality.

B Summary statistics

Table 5: Summary statistics for ETF-level analysis

Statistic	N	Mean	St. Dev.	Min	Q_{25}	Q_{75}	Max
$ret_{i,j,[-6,-4]}$	1,188	0.0003	0.007	-0.036	0.000	0.0004	0.184
$ret_{i,j,[-4,-2]}$	1,188	0.0003	0.005	-0.046	-0.001	0.002	0.065
$ret_{i,j,[-2,0]}$	1,188	0.0001	0.004	-0.030	-0.001	0.001	0.023
$ret_{i,j,[0,2]}$	1,188	0.00001	0.003	-0.026	-0.001	0.001	0.027
$ret_{i,j,[2,4]}$	1,188	-0.00002	0.005	-0.099	-0.001	0.001	0.042
$ret_{i,j,[4,6]}$	1,188	0.0002	0.012	-0.036	-0.0003	0.0005	0.375
SUR	1,188	-0.0003	0.024	-0.417	-0.002	0.002	0.496
$SUR_{1_{Weight<Q_{0.25}}}$	210	0.0005	0.038	-0.133	-0.002	0.002	0.496
$SUR_{1_{Weight\in[Q_{0.25},Q_{0.5}]}}$	279	-0.002	0.030	-0.417	-0.001	0.002	0.174
$SUR_{1_{Weight\in[Q_{0.5},Q_{0.75}]}}$	332	0.001	0.020	-0.150	-0.002	0.002	0.174
$SUR_{1_{Weight>Q_{0.75}}}$	367	-0.001	0.007	-0.042	-0.001	0.002	0.023

Note: The variables in the table are: ETF returns over specified announcement window, ret; measure of earnings surprise of a stock within a corresponding ETF, SUR; the indicator function that takes the value of 1 if the weight of stock in ETF is in the q^{th} quartile of ETF weights distribution, $I_{W\in q}$; dummy variable that takes the value 1 when the average news score or the number of relevant events on the U.S. market that day is larger than median, $BUSY$. The description of variables is in Section 3.1. The sample period is 2016-2017.

Table 6: Summary statistics for stock-level analysis

Statistic	N	Mean	St. Dev.	Min	Q_{25}	Q_{75}	Max
$aret_j$	684,115	0.0004	0.018	−0.880	−0.008	0.008	0.536
$aret_i$	684,115	−0.001	0.006	−0.065	−0.003	0.002	0.043
AIV_i	684,115	−0.186	0.708	−6.334	−0.604	0.271	2.647
AIV_j	684,086	−0.374	0.658	−5.074	−0.731	0.071	2.088
$\frac{1}{P}$	684,115	0.939	47.990	0.00000	0.024	0.197	9,021.200
$log(Mkt\ Cap)$	684,115	22.070	1.956	14.266	20.597	23.291	32.067
$ret_{[-1]}$	684,115	0.001	0.022	−1.361	−0.007	0.010	2.324
$ret_{[-3,-2]}$	684,106	−0.001	0.064	−2.453	−0.008	0.009	2.350
$ret_{[-6,-4]}$	684,064	0.002	0.055	−2.362	−0.011	0.014	2.345
$Amihud$	640,649	0.00000	0.00005	0.000	0.000	0.00000	0.016
$log(Amihud)$	634,130	−17.980	2.336	−49.451	−19.561	−16.374	−4.146
Δ	684,115	0.00003	0.003	0.000	0.000	0.000	0.850
W	684,115	0.285	1.935	0.000	0.007	0.104	99.400

Note: The variables in the table are: Fama-French adjusted cumulative abnormal returns of non-announcing stock, $aret_j$, and ETF, $aret_i$; the abnormal idiosyncratic volatility of ETF, AIV_i, and non-announcing stock in given ETF, AIV_j; the inverse of price of non-announcing stock, $\frac{1}{P}$; the log of market capitalization of non-announcing stock, $log(Mkt\ Cap)$; the lagged returns of non-announcing stock ($ret_{[-1]}$, $ret_{[-3,-2]}$, $ret_{[-6,-4]}$); Amihud illiquidity measure of non-announcing stock, $Amihud$, and the log of it, $log(Amihud)$; intensity of arbitrage, Δ; and the weight of non-announcing stock in ETF, W. The description of variables is in Section 3.2. The sample period is 2016-2017.

Table 7: Correlation between abnormal stock returns of non-announcing stocks and ETFs around earnings releases

	Dependent variable: Fama-French Adjusted Cumulative Returns			
	$aret_{[t-1,t]}$	$aret_{[t,t+1]}$	$aret_{[t,t+5]}$	$aret_{[t,t+10]}$
	Response around announcements in the first weight quartile: $W_{i,k} \in [Q_1, Q_2]$			
$aret_{ETF,t}$	0.004	0.075***	0.268**	-0.033
	(0.034)	(0.007)	(0.105)	(0.163)
$aret_{ETF,t} * Weight$	-0.008	-0.020	0.074	0.064
	(0.019)	(0.015)	(0.082)	(0.091)
Controls	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes
Observations	118,225	118,222	118,217	117,171
R^2	0.169	0.175	0.255	0.266
	Response around announcements in the second weight quartile: $W_{i,k} \in [Q_2, Q_3]$			
$aret_{ETF,t}$	-0.036	-0.191**	-0.139	0.465
	(0.084)	(0.093)	(0.133)	(0.447)
$aret_{ETF,t} * Weight$	-0.0004	0.009	-0.019	0.010
	(0.008)	(0.008)	(0.034)	(0.038)
Controls	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes
Observations	157,178	157,175	157,162	153,693
R^2	0.138	0.144	0.218	0.329
	Response around announcements in the third weight quartile: $W_{i,k} \in [Q_3, Q_4]$			
$aret_{ETF,t}$	-0.010	-0.069	0.023	0.131
	(0.057)	(0.093)	(0.127)	(0.209)
$aret_{ETF,t} * Weight$	-0.026	-0.014	-0.008	-0.043
	(0.026)	(0.029)	(0.036)	(0.037)
Controls	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes
Observations	177,872	177,863	177,847	175,346
R^2	0.143	0.160	0.160	0.172
	Response around announcements in the fourth weight quartile: $W_{i,k} > Q_4$			
$aret_{ETF,t}$	0.054	0.174	0.152	0.041
	(0.052)	(0.106)	(0.126)	(0.189)
$aret_{ETF,t} * Weight$	0.020***	0.002	-0.021	-0.022
	(0.006)	(0.010)	(0.016)	(0.022)
Controls	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes
Observations	180,788	180,778	180,772	167,075
R^2	0.157	0.166	0.139	0.199
Note:				*p<0.1; **p<0.05; ***p<0.01

Note: This table presents estimates of regression of the non-announcing stock cumulative abnormal return for a given period on the abnormal return of the ETF for different weight quartiles of announcing stocks. $Controls$ include the inverse of price of non-announcing stock, $\frac{1}{P}$, the log of market capitalization of non-announcing stock, $log(Mkt\ Cap)$, the log of Amihud illiquidity measure of non-announcing stock, $log(Amihud)$, and the lagged returns of non-announcing stock ($ret_{[-1]}$, $ret_{[-3,-2]}$, $ret_{[-6,-4]}$). We use fund, stock, and day fixed effects. Standard errors are clustered at the fund-stock and reported in parentheses. The description of variables is in Section 3.2. The sample period is 2016-2017.