Stock-Specific Price Discovery From ETFs*

Thomas Ernst†

March 22, 2022

Abstract

Conventional wisdom warns that exchange-traded funds (ETFs) “steal” single-stock liquidity or force co-movement. Contra this belief, I present a theoretical model with empirical evidence demonstrating investors with stock-specific information trade both single stocks and ETFs. This tandem trading allows single-stock investors access to ETF liquidity, and flexible price adjustments. Using trade-level data on over 500 ETFs, I exploit exchange latencies to show that investors place simultaneous, same-direction trades in both a stock and ETF. Consistent with my model, effects are strongest when an individual stock has a large ETF-weight or large informational asymmetry. I conclude ETFs provide single-stock price discovery.

*First Draft: October 2017. I am very grateful to my advisors: Haoxiang Zhu, Chester Spatt, Leonid Kogan, and Jiang Wang. Additional helpful comments were provided by Maureen O'Hara, Thierry Foucault, Albert Menkveld, Andrey Malenko, Simon Gervais, Shimon Kogan, Antoinette Schoar, Jonathan Parker, Lawrence Schmidt, David Thesmar, Hui Chen, Daniel Greenwald, Deborah Lucas, Dobrislav Dobrev, Andrew Lo, Robert Merton, Christopher Palmer, Adrien Verdelhan, Eben Lazarus, Austin Gerig, Peter Dixon, Paul Tetlock, and Eddy Hu, as well as seminar participants at MIT, NBER Big Data Conference, the Microstructure Exchange, McGill University, University of Maryland, University of Georgia, Boston College, Plato Partnership, and the Securities and Exchange Commission.

†University of Maryland, Robert H. Smith School of Business: ternst@umd.edu

I. Introduction

Do exchange-traded funds (ETFs) harm price discovery? With an average trading volume of \$90 billion per day, ETFs now comprise 30% of total US equities trading volume. This ascendancy raises two primary concerns about the impact of ETFs on the price discovery process. The first is that ETFs—with their promise of diversification, ease, and safety—lure noise traders away from individual stocks. Consequently, informed traders lose profits, and thus lose incentives to acquire stock-specific information. The second, related concern is that ETF trading could lead to excessive co-movement. Given the high volume of ETF trading, a liquidity shock in an ETF could trigger arbitrage activity that forces a symmetric price movement in all the underlying securities, regardless of stock fundamentals. Underlying both concerns is the implicit assumption that investors with stock-specific information either never trade ETFs, or that the two assets function as separate markets.

This paper is the first to theoretically model and empirically document the conditions under which investors with stock-specific information strategically trade both stocks and ETFs in tandem. This trading behavior attenuates the concerns mentioned in the previous paragraph: stock-specific investors can access ETF liquidity, and stock prices can flexibly adjust to ETF price movements. My model shows that ETF-based price discovery of single-stock information occurs whenever a stock is a sufficiently large or sufficiently volatile constituent of the ETF, or both. The predictions of the model are supported by a series of empirical tests using high frequency transactions data of single stocks and ETFs.

Formally, the model is an extension of Glosten and Milgrom (1985) with three assets: stock A, stock B, and an ETF AB which combines ϕ shares of A and $(1-\phi)$ shares of B. There is a single market maker who posts quotes for all three assets. Each asset has some level of noise trading. In the simplest formulation of the model, there are informed investors only in stock A while the price of B remains fixed. The investors with information about stock A (hereafter referred to as "A-informed" traders) know the value of A exactly, and can trade both the stock and the ETF, subject to their position limit. While trading the ETF gives less exposure to stock A per unit of capital, the ETF is available at a lower bid-ask spread. This position limit induces a strategic choice of where to trade: the more capital investors commit to the ETF, the less capital they can commit to the individual stock. I normalize the capital limit to a single share, and model the trade-both behavior of traders by randomization between pure strategies of trading either the stock or the ETF. Two cases of equilibria obtain. The first case is a separating equilibrium in which A-informed traders only trade A and the ETF is available with a zero bid-ask spread. The second case is a pooling equilibrium in which A-informed traders randomize between trading A and trading the ETF, and equilibrium spreads leave informed investors indifferent between trading A or the ETF. In the pooling equilibrium, A-informed traders profit from market noise traders in the ETF, and the market maker learns about the value of stock A following an ETF trade.

When the ETF weight and informational asymmetry in stock A are sufficiently high, the pooling equilibrium prevails. As an example, consider the Technology Sector SPDR ETF, which trades under the ticker XLK. The Technology SPDR is value-weighted; hence Apple, being a large firm, comprises 19% of the ETF. Investors with a modest Apple-specific informational advantage also have an information advantage about XLK. As a result, trading both securities can offer more profit than trading Apple alone. On the other hand, Paypal, being a much smaller firm, comprises only 2% of the Technology SPDR. The portion of price movements in XLK explained by Paypal, however, can be much larger if the volatility of Paypal increases relative to other stocks. At times when the volatility of Paypal rises to three or four times that of other technology stocks, investors with Paypal-specific information also have a substantial information advantage about XLK.

The full model also has a class of traders with private information about stock B (hereafter "B-informed"). Both A-informed and B-informed traders have freedom over which asset to trade. This yields two additional equilibria: a partial separating equilibrium and a fully pooling equilibrium. In the fully pooling equilibrium, both A-informed and B-informed traders randomize between their respective stock and the ETF. In the partial separating equilibrium, only one class of informed trader trades the ETF. In both of these equilibria, different pieces of stock-specific information act as substitutes. If A-informed traders send more orders to the ETF, the market maker has to increase the ETF bid-ask spread. This increased ETF spread reduces the profit B-informed traders could make trading the ETF, so they reduce the probability with which they trade the ETF or stop trading the ETF altogether. Traders with small pieces of information or information about small stocks can be "excluded" from trading the ETF whenever the value of their information is less than cost of the adverse selection they face when trading the ETF.

This conflict between traders with different pieces of information creates the screening power of the ETF. As an example, consider the XLY Consumer Discretionary ETF. The ETF is value weighted, so the large retailer Amazon has a weight of 22% while the very small retailer Gap, Inc. has a weight of just 0.27%. An investor with Gap-specific information who attempts to exploit their informational advantage by trading XLY is unlikely to profit if traders with Amazon-specific information are also trading XLY. As a result, the partial separating equilibrium prevails and investors informed about Gap do not trade XLY.

In the fully pooling equilibrium the market maker learns about both stocks following an ETF trade. Dynamics between ETF trades and stock quotes depend on more than just the weight of each stock in the ETF. The population of informed traders and the market maker's prior both regulate the influence of ETF price movements on the underlying stocks. Following an ETF trade, stocks with certain priors or few informed traders undergo small adjustments in price, whereas stocks with uncertain priors or many informed traders undergo large adjustments in price. While both stock quotes change following an ETF trade in the fully pooling equilibrium, this movement is a flexible process of adjustment.

I test the predictions of the model with NYSE TAQ data from August 1, 2015 to December 31, 2018. I analyze all stocks listed in the S&P 500 during this sample period, and 552 ETFs which meet the criteria of having at least 50% of their holdings invested in the aforementioned stocks and trade on at least ten days. Under the model, investors should trade both the stock and ETF whenever stock weight in the ETF is high and the stock-specific informational asymmetries are large. While TAQ data is anonymous, I exploit exchange latencies and precise timestamps to identify investors who trade both the stock and the ETF simultaneously. These simultaneous stock–ETF trades give a high resolution measure of how investors trade stocks and ETFs, and I examine how this relationship varies across stock-specific characteristics.

Consistent with the model predictions, I find that these simultaneous trades are driven by stock-specific information. When a stock has an earnings date, large absolute return, or stock-specific news article or press release published, that stock sees an increase in simultaneous stock–ETF trades. Effects are much stronger for large ETF-weight stocks than for small ETF-weight stocks, and are also much stronger for large absolute returns than for small absolute returns. When trades are signed according to Lee and Ready (1991), I find that the simultaneous trades are in the same direction: investors buy both the stock and the ETF at the same time, or they sell both at the same time. This same-direction trade-both behavior matches my model and is inconsistent with alternative explanations like hedging or market making.

Simultaneous trades are a sizable portion of trading volume. Simultaneous trades from a single stock–ETF pairing regularly comprises 1% to 2% of ETF total volume. Market makers appear to view these simultaneous trades as well informed, as they have larger-than-average price impacts. Simultaneous trades also earn negative realized spreads, consistent with placement by informed traders. A typical order pays a realized spread of one cent per share in stocks, and a fraction of a cent on ETFs. Simultaneous trades are the opposite, with simultaneous orders *earning* a realized spread of one cent per share on the stock side of the trade, and *earning* a half cent per share on the ETF side of the trade. This tandem trading limits harm to price discovery from ETFs: investors in large or volatile stocks can use their stock-specific information to trade ETFs, and market makers learn from these trades.

II. Relation to Prior Literature

Early work on index funds focuses on the information shielding offered by basket securities. Gorton and Pennacchi (1991) first analyzed how different pieces of private information average out in an index fund, and thus liquidity traders can avoid informed traders by trading index funds.[1] A second line of the literature considers arbitrage or the information linkages between ETFs and the underlying assets Malamud (2016).[2] Subrahmanyam (1991) analyzes how introducing a basket changes the profits of traders, under the assumption that risk-neutral informed investors always trade both stocks and ETFs. In this theoretical model, the ETF is informative about the idiosyncratic component of even the smallest stocks.

My model innovates on these foundational papers by giving investors a strategic choice between trading stocks and ETFs. I find conditions under which investors with stock-specific information trade stocks and ETFs in tandem, whereby prior concerns over ETFs are attenuated. Even if noise

[1] More recently, Bond and Garcia (2018) find an equilibrium of information acquisition while Cong and Xu (2016) model security design.

[2] Cespa and Foucault (2014) models the information linkage between two correlated assets, while Bhattacharya and O'Hara (2016) examines the linkage between an ETF and underlying under the assumption that stock-specific traders cannot trade the ETF. Chinco and Fos (2021) show how ETF rebalancing can create noise trading.

traders move to ETFs, I find that informed traders can follow them provided that the ETF weight or volatility of the stock is not too small. When the ETF weight and volatility of a stock are instead both small, however, I find that informed traders in that stock only trade the single stock. My model also combines this strategic choice with a single market maker across all securities, so that when ETF trades occur, market makers acquire stock-specific information, and update beliefs accordingly.

ETFs as a basket security can be thought of as a derivative, where the ETF price depends on the price of the underlying basket. While the ETF appears to be redundant, Back (1993) demonstrates how these assets can become informationally non-redundant in the presence of asymmetric information. In my model, the ETF gives the same terminal payoff as a weighted average of the underlying stocks, but a trade in the ETF conveys different information, and therefore equilibrium bid and ask prices in the ETF differ from those of the underlying securities. This difference reflects the differences in asymmetric information rather than arbitrage.[3] Box, Davis, Evans, and Lynch (2021) provide empirical support for this feature of ETFs, as they show that ETF prices adjust via passive quote updates rather than active trades.

ETFs are an important venue for price discovery. The empirical literature begins with Hasbrouck (2003), which compares price discovery in ETF markets with price discovery in futures markets and breaks down the share of price innovations that occur in each market.[4] The ETF-asset relationship has been extensively studied, with ETFs being associated with greater co-movement[5],

[3]The ETF acts an additive composition of the underlying stocks. As a result, strategic traders view different pieces of stock-specific information as substitutes, similar to non-ETF models of Foster and Viswanathan (1996) and Back, Cao, and Willard (2000) and incontrast to implement-complementarity models like Goldstein, Li, and Yang (2013).

[4]Yu (2005) compare the ETF and cash market, while Madhavan, Sobczyk, and Ang (2019) performs a spectral decomposition of ETF components. Cespa and Foucault (2014) investigate information spillovers between the SPY, E-Mini, and S&P 500 during the flash crash of May 6, 2010.

[5]Israeli, Lee, and Sridharan (2017) look at the fraction of a company's shares owned by ETFs, and find that when the fraction increases, the stock price begins to co-move more with factor news and co-move less with stock-specific fundamentals. Ben-David, Franzoni, and Moussawi (2018) argue ETFs can increase volatility and lead stocks to co-move beyond their fundamentals. Hamm (2014) shows the co-movement with ETFs is strong only in companies with low-quality earnings.

changes in stock-specific liquidity[6], and changes in price discovery.[7]

A key innovation of my paper is to examine relationship between the ETF and a single underlying asset. This relationship is not governed by arbitrage, as the law of one price can only compare the ETF with the entire basket of underlying assets. As a result, the conditions for Hasbrouck (1995) are not met. To surmount this challenge, I develop empirical methods of my paper built on the analysis outlined in Dobrev and Schaumburg (2017), who use trade time-stamps to identify cross-market activity. I utilize exchange-reported gateway-to-trade-processor latencies to identify simultaneous trades; a different approach is developed by Aquilina, Budish, and O'Neill (2020) to analyze same-asset simultaneous messages. In my setting, I document simultaneous stock-ETF trades which are in the same direction, highly profitable, and driven by stock-specific information. The use of this novel empirical technique allows me to establish differences between the large-stock–ETF relationship and small-stock–ETF relationship. Traders with information about large stocks or large informational asymmetries can and do profitably trade the ETF, while traders with information about small stocks or small information asymmetries cannot.

Price discovery can happen across multiple assets or venues. Easley and O'Hara (1987) model how traders use both upstairs and downstairs markets. Johnson and So (2012) study how informed traders use options as well as stocks. Holden, Mao, and Nam (2018) show how price discovery happens across both the stock and bonds of a company. Hasbrouck (2018) estimates the informational contribution of each exchange in equities trading. My paper demonstrates that investors with stock specific information trade both stocks and ETFs; as a result, ETF trades contribute to the price discovery of stock-specific information.

[6] Sağlam, Tuzun, and Wermers (2019) present evidence from a difference-in-difference estimation which shows higher ETF ownership leads to improved liquidity for the underlying stocks under normal market conditions, though the effects may be reversed during periods of market stress. Huang, O'Hara, and Zhong (2021) collect evidence that suggests industry ETFs allow investors to hedge risks especially when shorting the ETF, while Lu and Qin (2021) shows price effects from levered ETFs. Brogaard, Heath, and Huang (2019) present evidence that ETF index-tracking trades increase liquidity for stocks which are selected for creation/redemption baskets, and harm liquidity for stocks omitted from this basket.Dannhauser (2017) shows that bonds included in ETFs have higher prices, but decreased liquidity trader participation, while Dannhauser and Hoseinzade (2022) and Haddad, Moreira, and Muir (2021) examine bond ETFs during the covid crisis.

[7] Glosten, Nallareddy, and Zou (2016) show that ETFs allow more efficient incorporation of factor-based information in ETFs. Easley, Michayluk, O'Hara, and Putniņš (2021) show that many ETFs go beyond tracking the broad market, and instead offer portfolios on specific factors, while Bae and Kim (2020) show illiquid ETFs have large tracking errors. Bessembinder, Spatt, and Venkataraman (2019) suggest that ETFs could help bond dealers hedge inventory risks. With trade and inventory data, Pan and Zeng (2016) confirm this. Holden and Nam (2019) find that ETFs lead to liquidity improvements in illiquid bonds. Evans, Moussawi, Pagano, and Sedunov (2019) suggest ETF shorting by liquidity providers improves price discovery.

III. The Model

A. Assets

The model is in the style of Glosten and Milgrom (1985), with two stocks, A and B. Each stock in the economy pays a single per-share liquidating dividend from $\{0, 1\}$, and I assume the two dividends are independent. One share of the market portfolio contains ϕ shares of stock A and $(1 - \phi)$ shares of stock B.

The economy also has an ETF, which has the same weights as the market portfolio. Thus each share of the ETF contains ϕ shares of stock A and $(1 - \phi)$ shares of stock B. With market weights, no rebalancing is needed: should the value of stock A increase, the value of the ϕ shares of stock A within the ETF increases. I compare this economy with the alternative no-ETF economy where only stocks A and B can be traded in Section IV C.

B. Market Maker

There is a single competitive market maker who posts quotes in all three securities. The market maker is risk neutral and has observable Bayesian prior beliefs:

$$\mathbb{P}(A = 1) = \delta, \qquad \mathbb{P}(B = 1) = \beta$$

In each security, the market maker sets an ask (or bid) price equal to the expected value of the security conditional on receiving an order to buy (or sell). This expected value depends on both the population of traders and their trading strategies. Following an order arrival, the market maker updates beliefs about security value. Traders arrive according to a Poisson process, and the unit mass of traders can be divided up into informed and uninformed traders. Figure 1 presents an overview of the model timing.

C. Uninformed Traders

Uninformed traders trade to meet inventory shocks from an unmodeled source. These uninformed, or noise, traders can be divided into three groups based on the type of shock they receive:

- Stock A noise traders of mass σ_A

Figure 1. Timeline of the Model. A single risk-neutral, competitive market maker posts quotes in all three securities. A single trader arrives and trades against one of these quotes. Following an order, the market maker updates beliefs about the value of each of the three securities.

- Stock B noise traders of mass σ_B
- Market-shock noise traders of mass σ_M

Noise traders buy or sell the asset with equal probability. I also normalize the demand or supply of each noise trader to be a single share of the asset they choose to trade.[8] Uninformed traders who experience a stock-specific shock trade only an individual stock. Uninformed traders who experience the market shock could satisfy their trading needs with either the ETF or a combination of stocks A and B. Trading the ETF, however, allows market-based noise traders to achieve the same payoff at a potentially lower transaction cost. This possibility arises because the ETF offers some screening power. Trading at the ETF quote gives an investor the ability to buy or sell stocks A and B in a fixed ratio. Trading at the individual stock quotes, by contrast, allows an investor the ability to trade any ratio of stocks A and B. Thus for any information structure, the ETF quotes must be at least as good as a weighted sum of the individual quotes.[9] In all equilibria of my model, the ETF quotes turn out to be strictly better than the weighted sum of the individual quotes.

D. Informed Traders

Informed traders know the value of exactly one of the two securities. There is a mass μ_A of traders who know the value of stock A, and a mass μ_B of traders who know the value of stock B. Informed traders are limited to trading only a single unit of any asset, though they can randomize

[8]The model could be extended as in Easley and O'Hara (1987) to have noise traders trading multiple quantities or multiple assets.

[9]Reversing this result requires a non-information friction. For example, adding factor-informed traders as in Appendix B does not reverse the result as it is a purely informational friction. One natural case of a non-informational friction would be to add a fixed-cost trading friction so that investors would be willing to trade the ETF even if it had a wider information-based spread.

their selection. Appendix B provides a formal definition of factor risk in the model, and also extends the model to include the presence of factor-informed traders. Intuitively, the single-unit limitation on trade captures a cost of capital or risk limit for their trading strategy. While investors could trade more aggressively in the ETF to obtain the same stock-specific exposure, this would require significantly higher capital or exposure to significantly more factor risk. Informed investors therefore face the following tradeoff: they can choose to trade A at a wide spread, or they can choose to trade the ETF (AB) at a narrow spread with the caveat that the ETF contains only $\phi < 1$ shares of A.

In addition to being conceptualized as the ETF weight of each stock, ϕ can be thought of in more general terms as the relevance of the investor's information. When an investor has information about security A, they could also trade a closely related security (AB). While the investor's information is less relevant to the price of (AB), the asset may be available at a lower trading cost. The lower ϕ, the less relevant the information, and thus the less appealing committing capital to this alternative investment becomes.

E. *Equilibrium*

The Bayesian-Nash equilibria between the traders and the market maker obtains as follows. Let A-informed traders μ_A submit orders to the ETF with probability ψ_A and B-informed traders μ_B submit orders to the ETF with probability ψ_B. A pair of strategies (ψ_A, ψ_B) is an equilibrium strategy when, for each stock, either ψ_i leaves the informed trader indifferent between trading the stock and the ETF, or when $\psi_i = 0$ and the informed trader strictly prefers to trade the single stock.

The effectiveness of the ETF for screening informed traders varies across the different equilibria. In a fully separating equilibrium, where investors with stock-specific information only trade specific stocks, there is no adverse selection in the ETF. ETF screening of stock-specific information is rarely complete, however. In a fully pooling equilibrium, traders from both stock A and stock B trade the ETF, while in a partial separating equilibrium, one class of informed investors trades both the stock and the ETF.[10]

[10]Easley and O'Hara (1987) develop the use of separating and pooling equilibrium in a single asset with multiple possible trading volumes. This paper uses a similar approach to model the possibility of trading on the same information in multiple assets. The novel partial-separating equilibrium arises because there are two classes of

The direct modeling of bid-ask spreads, while allowing ETFs and underlying stocks to have different levels of adverse selection, means there are no law of one price violations. The word arbitrage is often used in reference to ETFs. In these industry applications of the term arbitrage, however, there are no true law of one price violations.[11] Consistent with observed behavior of ETF prices, the model has no law of once price violations.[12]

IV. Price Discovery in a Single Asset

In this section, I analyze the equilibria that result with price discovery in only stock A. Thus I set $\mu_B = 0 = \sigma_B$ and $\beta = \frac{1}{2}$. In this simplified setting, there are only two possible equilibria. The first is a separating equilibrium, where investors with information about A trade only security A and do not the ETF. The second is a pooling equilibrium, where investors with information about A mix their orders, randomizing their order to either A or the ETF. In this equilibrium, investors are indifferent between trading the single stock and the ETF, as the profit from trading the single stock at a wide spread is the same as the profit from trading the ETF at a narrow spread.[13] Figure 2 illustrates the sequence of possible trades while Table I reviews key model parameters.

informed traders: those informed about A and those informed about B.

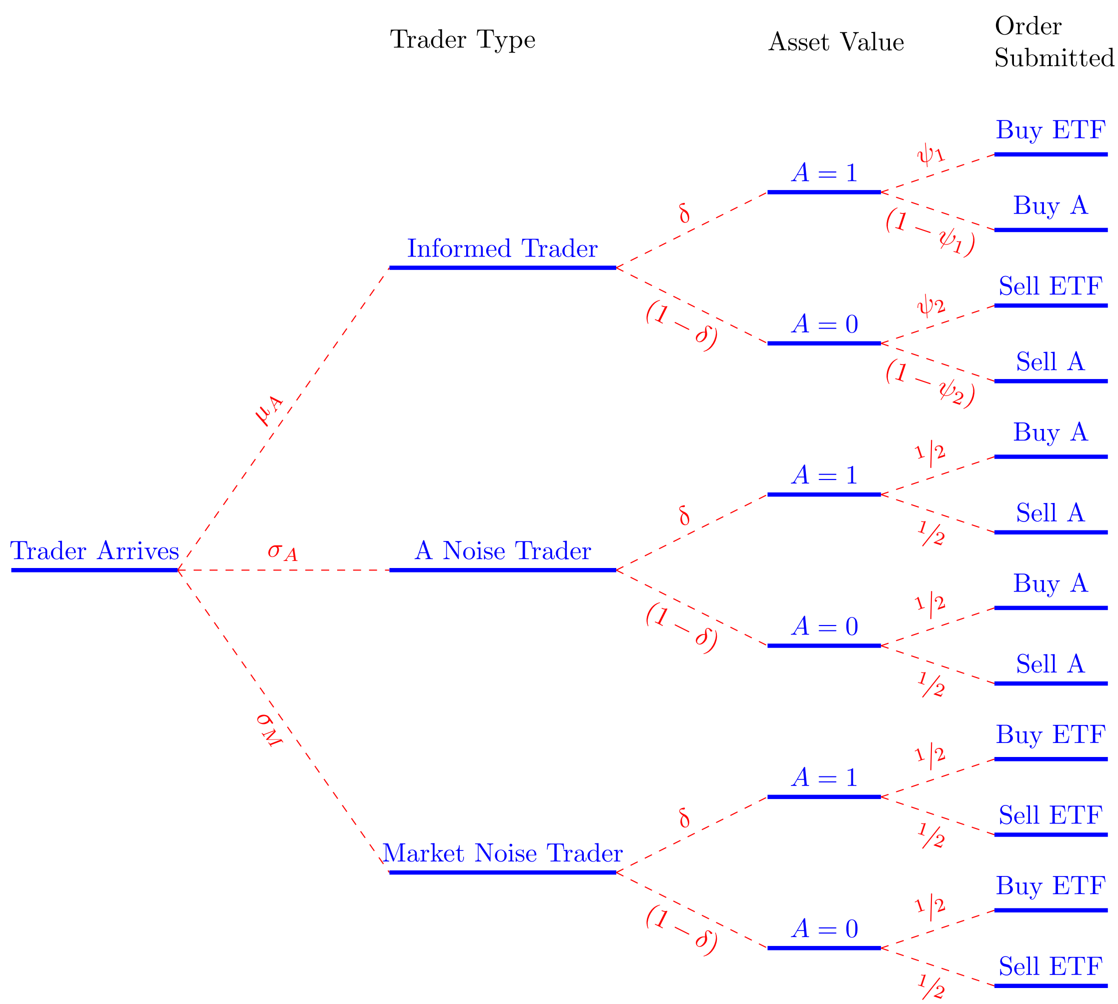
[11]The creation/redemption mechanism is sometimes referred to as an "arbitrage" mechanism. After the close of the market, authorized participants (APs) can exchange underlying baskets of securities for ETF shares. The securities exchanged, however, had to have been acquired during trading hours. Positions in securities could be acquired for a variety of reasons, including regular market-making activities, so the use of the creation/redemption mechanism does not imply any previous violation of the law of one price.

Deviations from intraday net asset value (iNAV) are also sometimes referred to as "arbitrage" opportunities. They typically arise, however, from the technical details of the iNAV calculation, as discussed in Donohue (2012). iNAV is usually calculated from last prices of the components, so a deviation from iNAV is typically staleness in prices; Madhavan and Sobczyk (2016) provide documentation of this phenomenon . iNAV can also be computed from bid prices; in this case, iNAV just confirms that the risk from placing one limit order for the ETF can differ from the risk of placing many limit orders on each of the basket securities. For some securities, the creation/redemption basket is different from the current ETF portfolio, so iNAV, which reflects the creation/redemption basket, can differ from the market price of the current portfolio. Finally, errors are common in the calculation and reporting of iNAV values.

[12]KCG analysis on trading for the entire universe of US equity ETFs finds that arbitrage opportunities occur in less than 10% of ETFs. These arbitrages occurred in smaller, much less liquid ETFs, and were always less than $5,000, which is "unlikely enough to cover all the trading, settlement, and creation costs." Mackintosh (2014) Box et al. (2021) confirm that when stock prices move, ETF spreads tend to be wider, and quotes are adjusted without arbitrage trades.

[13]Note that it is not possible for an equilibrium to obtain where A-informed investors only trade the ETF. If this were the case, then security A would have no spread, and the A-informed investors would earn greater profits trading security A, as they do in the separating equilibrium.

Figure 2. Sequence of Potential Orders. Traders can be Stock A noise traders, market noise traders, or informed traders. Stock A noise traders buy or sell stock A while market noise traders buy or sell the ETF. In the separating equilibrium, informed traders maximize their profits by trading only Stock A, and not the ETF (i.e. $\psi = 0$). In the pooling equilibrium, informed traders randomize between trading the stock and the ETF. The equilibrium probability ψ of trading the ETF, induces the market maker to quote a spread which leaves informed traders indifferent between trading stock A at a wide spread or trading the ETF at a narrow spread.



A. Separating Equilibrium

PROPOSITION 1: *A separating equilibrium in which informed traders only trade A and do not trade the ETF, obtains if and only if:*

$$\phi \leq \frac{\frac{1}{2}\sigma_A}{(1-\delta)\mu_A + \frac{1}{2}\sigma_A} \qquad \textit{(bid condition)}$$

$$\phi \leq \frac{\frac{1}{2}\sigma_A}{\delta\mu_A + \frac{1}{2}\sigma_A} \qquad \textit{(ask condition)}$$

In the separating equilibrium, traders with information about security A only submit orders to stock A, and do not trade the ETF. Since no informed orders are submitted to the ETF, there is no information asymmetry and orders in the ETF reveal no information about the underlying value of the assets. Therefore, the ETF is offered at a zero bid-ask spread.

For the separating equilibrium to hold, the payoff to an informed trader from trading the individual stock must be greater than the payoff from trading the ETF at no spread:

$$\text{Payoff from trading ETF} \leq \text{Payoff from trading stock}$$

$$\phi(1-\delta) \leq 1 - ask_A \tag{1}$$

$$\phi\delta \leq bid_A \tag{2}$$

For the market maker to make zero expected profits, each quote must be the expected value of A conditional on receiving a market order. The asking price is the expected value of A conditional on receiving a buy order in A. Since stock A pays a liquidating dividend from $\{0, 1\}$, the expected value of stock A is just the probability that A pays a dividend of 1. A similar logic holds for the bid price. The bid and ask are therefore given by:

$$\begin{aligned} ask = \mathbb{P}(A = 1|\text{buy}_A, \delta) &= \frac{\mathbb{P}(A = 1\&\text{buy}_A|\delta)}{\mathbb{P}(\text{buy}_A|\delta)} \\ &= \delta\frac{\mu_A + \frac{1}{2}\sigma_A}{\delta\mu_A + \frac{1}{2}\sigma_A} \\ bid = \mathbb{P}(A = 1|\text{sell}_A, \delta) &= \frac{\mathbb{P}(A = 1\&\text{sell}_A|\delta)}{\mathbb{P}(\text{sell}_A|\delta)} \\ &= \delta\frac{\frac{1}{2}\sigma_A}{(1-\delta)\mu_A + \frac{1}{2}\sigma_A} \end{aligned}$$

When spreads in the single stock become wide enough, either of (1) or (2) may no longer be satisfied. In this case, informed investors could make more profit trading the ETF than trading the individual stock. If they switched and only traded the ETF, then the single stock would have no spread, and trading the single stock would be more profitable. Thus for a non-separating equilibrium, informed traders must randomize between trading the ETF and the single stock.

B. *Pooling Equilibrium*

In the pooling equilibrium, informed traders randomize between trading stock A and trading the ETF. Figure 2 presents the possible trades for the pooling equilibrium. Informed traders trade the ETF with probability ψ and the stock with probability $(1-\psi)$. The equilibrium value of ψ leaves informed traders indifferent between trading either the stock or the ETF.

PROPOSITION 2: *A pooling equilibrium in which informed traders trade both Stock A and the ETF exists so long as either of the following conditions hold:*

$$\phi > \frac{\frac{1}{2}\sigma_A}{(1-\delta)\mu_A + \frac{1}{2}\sigma_A} \qquad \text{(bid condition)} \tag{3}$$

$$\phi > \frac{\frac{1}{2}\sigma_A}{\delta\mu_A + \frac{1}{2}\sigma_A} \qquad \text{(ask condition)} \tag{4}$$

In a pooling equilibrium, informed investors mix between A and the ETF and submit orders to the ETF with the following probability:

$$\textit{ETF Buy Probability (A=1):}\ \psi_1 = \frac{\phi\delta\mu_A\sigma_M - \frac{1}{2}(1-\phi)\sigma_A\sigma_M}{\delta\mu_A[\sigma_A + \phi\sigma_M]}$$

$$\textit{ETF Sell Probability (A=0):}\ \psi_2 = \frac{\phi(1-\delta)\mu_A\sigma_M - \frac{1}{2}(1-\phi)\sigma_A\sigma_M}{(1-\delta)\mu_A[\sigma_A + \phi\sigma_M]}$$

Note that these conditions are determined independently. For example, there could be a pooling equilibrium for bid quotes while the ask quotes have a separating equilibrium. This asymmetry can occur when the market maker's prior δ is far from $\frac{1}{2}$, and therefore the market maker views either $\mathbb{P}(A=1)$ or $\mathbb{P}(A=0)$ as more likely.

In the pooling equilibrium, an informed trader with a signal about A randomizes between trading the ETF and trading the single stock. While the trader obtains fewer shares of A by trading the ETF, the ETF has a much narrower spread. Once it is more profitable for an informed trader to randomize between the stock and ETF, the market maker must charge a spread for ETF orders. The market noise traders in the ETF must pay this spread when they trade, and thus pay some of the costs of the stock-specific adverse selection.

For the informed trader to be willing to use a mixed strategy, he must be indifferent between buying the ETF or buying the individual stock. In the single stock, he trades one share of A at the single-stock spread. In the ETF, he obtains only ϕ shares of A, but at the narrower ETF spread. Spreads in both markets depend on the probability ψ that he trades the ETF. Shifting more orders to the ETF increases the ETF spread and decreases the single-stock spread. In equilibrium, ψ_1 (proportion of informed buy orders sent to the ETF) and ψ_2 (proportion of informed buy orders sent to the ETF) must solve:

$$[\phi \cdot 1 + (1-\phi) \cdot (\frac{1}{2} - ask_{(AB)})] = 1 - ask_A$$
$$bid_{(AB)} - (1-\phi) \cdot \frac{1}{2} = bid_A$$

Solving for these expressions yields the mixing probabilities given in Proposition 2. A summary of the results is given in Table I. Note that for any $\phi > 0$, there exists a proportion μ of informed traders and a proportion σ_M of noise traders in the ETF such that a pooling equilibrium exists.

COROLLARY 1: *The portion of orders submitted to the ETF by A-informed traders is increasing in:*

1. *The number of informed traders,* μ_A.
2. *The number of noise traders in the ETF,* σ_M.
3. *The accuracy of the market maker's belief,* $-|A-\delta|$.
4. *The ETF weight of the stock,* ϕ.

The first three parameters determine the relative sizes of bid-ask spreads. If the are more informed traders, bid-ask spreads in stock A are wide. If there are more noise traders in the ETF, ETF spreads are narrower for a given level of informed trading. The accuracy of the market maker's belief, $|\delta - A|$, reflects both the sensitivity of the market maker's belief to order flow and

Table I: Key Variables and Equilibrium Spreads. This table summarizes the equilibrium spreads for the separating and pooling equilibria. In the separating equilibrium, informed investors only trade stock A; the ETF has no informed trading and thus no bid-ask spread. In the pooling equilibrium, informed investors randomize their orders, trading the ETF with probability ψ and the stock with probability $(1-\psi)$.

Key Model Parameters

Parameter	Definition
ϕ	Weighting of A in the ETF
δ	Market Maker's prior about $P(A=1)$
μ_A	Fraction of traders who are informed about A
σ_A	Fraction of noise traders with stock-A specific shock (thus trade stock A)
σ_M	Fraction of noise traders with market shocks (thus trade the ETF)
ψ_1	Fraction of informed traders who trade the ETF when $A=1$
ψ_2	Fraction of informed traders who trade the ETF when $A=0$

Separating Equilibrium Spreads

Security	Bid-Ask Quotes
A	$A_{bid} = \delta \frac{\frac{1}{2}\sigma_A}{(1-\delta)\mu_A + \frac{1}{2}\sigma_A}$
	$A_{ask} = \delta \frac{\mu_A + \frac{1}{2}\sigma_A}{\delta\mu_A + \frac{1}{2}\sigma_A}$
ETF	$(AB)_{bid} = \phi\delta + (1-\phi)\frac{1}{2}$
	$(AB)_{ask} = \phi\delta + (1-\phi)\frac{1}{2}$

Pooling Equilibrium Spreads

Security	Bid-Ask Quotes
A	$A_{bid} = \delta \frac{\frac{1}{2}\sigma_A}{(1-\delta)(1-\psi_2)\mu_A + \frac{1}{2}\sigma_A}$
	$A_{ask} = \delta \frac{(1-\psi_1)\mu_A + \frac{1}{2}\sigma_A}{\delta(1-\psi_1)\mu_A + \frac{1}{2}\sigma_A}$
ETF	$(AB)_{bid} = \phi\delta \frac{\frac{1}{2}\sigma_M}{(1-\delta)\psi_2\mu_A + \frac{1}{2}\sigma_M} + (1-\phi)\frac{1}{2}$ where $\psi_2 = \frac{\phi(1-\delta)\mu_A\sigma_M - \frac{1}{2}(1-\phi)\sigma_A\sigma_M}{(1-\delta)\mu_A[\sigma_A + \phi\sigma_M]}$
	$(AB)_{ask} = \phi\delta \frac{\psi_1\mu_A + \frac{1}{2}\sigma_M}{\delta\psi_1\mu_A + \frac{1}{2}\sigma_M} + (1-\phi)\frac{1}{2}$ where $\psi_1 = \frac{\phi\delta\mu_A\sigma_M - \frac{1}{2}(1-\phi)\sigma_A\sigma_M}{\delta\mu_A[\sigma_A + \phi\sigma_M]}$

the potential profits from trading. Suppose, for instance, that $A = 0$. If the market maker believes the value of A is close to zero, i.e. $\delta \sim 0$, then the market maker expects informed investors to sell. The bid price is very close to zero, leaving informed traders with little potential profit. As a result, the probability ψ with which they trade the ETF is very high. On the other hand, if the market maker believes $\delta \sim 1$, then the market maker does not expect informed traders to sell. The bid price is close to one, so informed traders prefer to trade the single stock. Trading the ETF is undesirable because the reduction in spread is small relative to the reduction in shares of A purchased.

The weight of the stock in an ETF, given by ϕ, determines the potential profit from using a mixed strategy. Investors with information about a large stock find themselves better informed about the ETF than they would with information about a smaller stock. The more informed a trader is about the ETF, the greater the profits they can make by trading against noise traders in the ETF.

Together, the weighting and spread create two dimensions along which stock–ETF interaction can vary. For most of the heavily traded ETFs, stock weights are determined by value weighting. Comparing stocks with a large ETF-weight against stocks with a small ETF-weight is therefore the same as comparing large market capitalization stocks with small market capitalization stocks. As Proposition 2 shows, however, the way in which stocks interact with the ETF also depends on informational asymmetries. For any fixed stock weight $\phi > 0$, there exists a pair of parameters (δ, μ_A) for which pooling is an equilibrium. When there are multiple informational asymmetries, this is no longer true. Section V explores how different pieces of information act as substitutes, and as a result investors with one piece of information may find that the ETF spread is always too wide for them to profit from trading the ETF.

C. *Effect of ETFs on Incentives to Acquire Information*

In this subsection, I examine the effect an ETF on the trading profits of informed traders. The impact of the ETF depends crucially on the viability of tandem stock-ETF trading strategies. When investors do not trade the ETF, as in the separating equilibrium, the impact of ETFs is unambiguously negative. Just as in Gorton and Pennacchi (1991), the ETF allows noise traders with market shocks to avoid trading single stocks and avoid losses to informed traders. This separating equilibrium, however, fails to exist for stocks with large ETF weight or large informational asymmetries. Instead, the pooling equilibrium prevails where informed traders profitably trade the ETF based on their stock-specific information. Proposition 3 compares investor profits in the presence and absence of ETFs. Profits are generally lower overall, but the reduction in profits is limited. As Corollary 1 demonstrates, the same parameters that allow for a liquid ETF allow for stock-specific informed trading in the ETF, and thus the reduction in profits from trading single stocks is balanced by improved profits from trading an ETF on stock-specific information.

PROPOSITION 3: *Let ζ_A denote the level of stock A noise traders in the absence of the ETF. Stock A informed trader profits are lower in the presence of the ETF so long as ζ_A satisfies:*

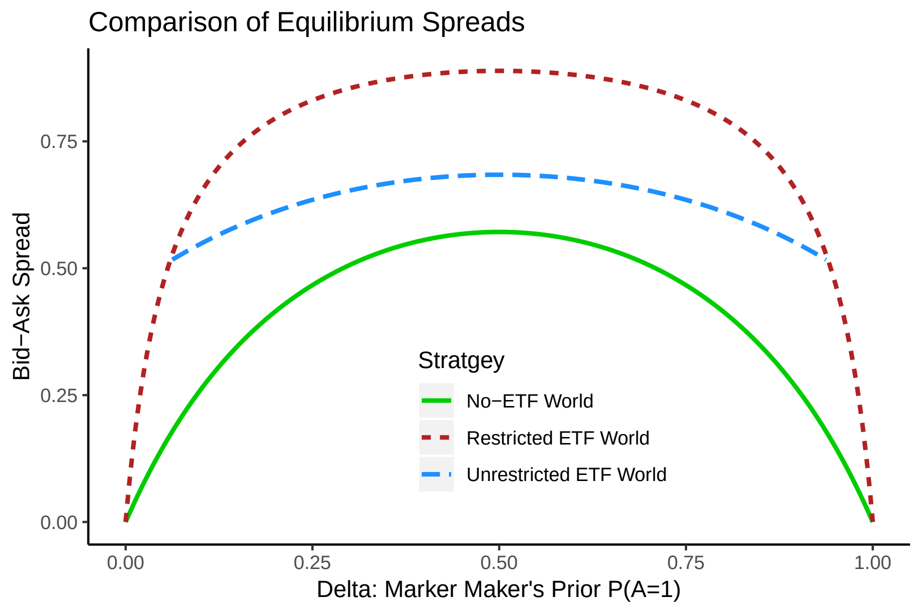
$$\zeta_A \leq \frac{\mu_A(1-\delta)[\sigma_A + \phi\sigma_M]}{\mu_A(1-\delta) - \frac{1}{2}(1-\phi)\sigma_M}$$

Profits for informed traders can only increase if $\zeta_A > \sigma_A + \phi\sigma_M$. This case is a natural one when noise traders adjust their volume with market conditions. Under Admati and Pfleiderer (1988), noise traders seek to trade at times when adverse selection is low. If traders view the ETF as providing some insulation against adverse selection, timing their trades may become less important. For the remainder of the paper, however, I will maintain the assumption that $\zeta_A = \sigma_A + \phi\sigma_M$. That is to say, the ETF does not change the overall level of noise trading, but only attracts existing noise traders with market shocks who would trade individual stocks in the absence of the ETF.[14] Under this assumption, the ETF unambiguously decreases informed profits, but the reduction in profits is limited in large stocks and in stocks with large informational asymmetries.

With the introduction of the ETF, stocks lose market noise trading volume in proportion to their market weights, with larger stocks losing a larger volume of traders. For informed investors in large stocks, however, their stock-specific information can also give them a serious informational advantage about the value of the ETF. For example, Chevron comprises 22% of the holdings of the Energy Select Sector SPDR ETF. Investors with an information advantage in Chevron cannot

[14] I assume constant σ_M as a fair common reference. The basic intuition of the paper is unchanged under the assumption that $\zeta_A \neq \sigma_A + \phi\sigma_M$. Bond and Garcia (2018) argue that larger stock-specific spreads will cause more stock-specific noise traders to decide to only hedge market risk, and become market risk noise traders. In contrast, Chinco and Fos (2021) propose a model where ETF-based rebalancing translates into noise trading in individual securities. Thus if σ_M traders leave for the ETF, and this ETF is one which requires active rebalancing, this will create some fraction ϵ of noise trading will be created the original asset.

Figure 3. Comparison of Equilibrium Spreads in Stock A. In the absence of the ETF, market noise traders trade individual stocks. This leads to an abundance of noise traders in stock A, and a narrow bid-ask spread (green line). With the introduction of the ETF, market noise traders switch to trading the ETF. If, as in previous literature, ETFs are assumed to totally screen out informed traders, stock A has wide spreads (red line). When informed traders are allowed to trade both stocks and ETFs, they do so. The resulting pooling equlibrium leads to improved spreads (blue line). Spread improvement from the trade-both strategy is most important when spreads are widest—at $\delta = .5$. Parameters: $\phi = .5, \sigma_M = .5, \mu_A = .4, \mu_B = 0, \sigma_A = \sigma_B = .05$.



help but have an information advantage in the Energy Sector ETF, and can take advantage of the low spreads and large depth of the ETF.

The benefit of trading the stock and ETF in tandem is largest precisely when stock-specific spreads are widest. Investors in the model have a position limit, so trading the ETF is not free. If investors trade the ETF with higher probability, they must trade the stock with lower probability. As Corollary 1 notes, the wider spreads are, the larger the portion of the trades investors send to the ETF. Figure 3 illustrates equilibrium spreads as a function of the market maker's prior belief δ that $A = 1$. Without an ETF, market noise traders must trade individual stocks, and spreads are narrow. Under the assumption from Gorton and Pennacchi (1991) that stock-specific informed investors do not trade ETFs, adding the ETF creates very wide spreads in the underlying stocks. As Figure 3 depicts, when informed traders are able to trade both securities, equilibrium spreads are closer to their values in the no-ETF world. The reduction in spreads brought about by the tandem trading strategy is largest precisely when the stock-specific spreads are widest—at $\delta = \frac{1}{2}$.

By trading both the stock and ETF, informed investors' profits, and equilibrium spreads, remain very similar between the world with ETFs, and one without.

V. Price Discovery with Multiple Assets

To fully develop the model, I now add μ_B traders who are informed about the value of stock B.[15] Stock B pays a liquidating dividend from $\{0,1\}$, and the market maker has a prior belief $P(B=1)=\beta$. I also assume that security B is uncorrelated with security A. Both classes of informed traders have a choice to trade one share of any of the securities. Given their stock-specific knowledge, A-informed investors consider trading stock A and the ETF (AB), while B-informed investors consider trading stock B and the ETF (AB). As before, investors can only trade a single share of any security, but they are allowed to randomize their selection.

Informed investors can now face adverse selection in the ETF. Each class of informed investors has information about only one stock; trading the ETF can expose them to adverse selection from the other stock in the ETF. There are now four potential equilibria. The first is a fully separating equilibrium, in which no informed traders submit orders to the ETF. For this equilibrium, the cutoffs are the same as in the previous section. The second is a fully pooling equilibrium, where both traders in A and B mix between the underlying security and the ETF. The last two equilibria are partial separating, where investors from one security randomize between trading the ETF and their single stock while investors from the other security only trade their single stock.

A. Partial Separating Equilibrium

Without loss of generality, I examine the partial separating equilibrium where A-informed traders randomize and trade both A and the ETF (AB), while B-informed traders only trade stock B. In this equilibrium, A-informed traders behave just as they did in Section IV B. The market maker must charge a non-zero bid-ask spread in the ETF to cover the costs of adverse selection from the A-informed traders. When the B-informed traders consider trading the ETF, they must now account for this non-zero bid-ask spread. For B-informed traders, the cutoff between using a

[15] I consider the case of factor-informed traders, who know the value of some underlying factor common to all securities, in Appendix C. The case is strikingly similar to the equilibria with informed traders in two independent stock-specific securities: factor traders trade the ETF, and there is the same potential for partial separating or fully pooling equilibria.

pure strategy of just trading stock B and using a mixed strategy of trading both B and the ETF is higher than it would be in the absence of the adverse selection from A-informed traders in the ETF.

PROPOSITION 4: *Partial Separating Equilibrium. If A traders mix between A and the ETF, B traders stay out of the ETF so long as:*

$$\text{If } B=0\text{: } \phi \geq \frac{\beta\left(\frac{(1-\beta)\mu_B}{(1-\beta)\mu_B+\frac{1}{2}\sigma_B}\right)\left((1-\delta)\mu_A+\frac{1}{2}\sigma_A+\frac{1}{2}\sigma_M\right)-\frac{1}{2}\delta\sigma_A}{\delta\left((1-\delta)\mu_A+\frac{1}{2}\sigma_A\right)+\beta\left((1-\delta)\mu_A+\frac{1}{2}\sigma_A+\frac{1}{2}\sigma_M\right)} \tag{5}$$

$$\text{If } B=1\text{: } \phi \geq \frac{(1-\beta)\left(\frac{\beta\mu_B}{\beta\mu_B+\frac{1}{2}\sigma_B}\right)\left(\delta\mu_A+\frac{1}{2}\sigma_A+\frac{1}{2}\sigma_M\right)-(1-\delta)\frac{1}{2}\sigma_A}{(1-\delta)\left(\delta\mu_A+\frac{1}{2}\sigma_A\right)+(1-\beta)\left(\delta\mu_A+\frac{1}{2}\sigma_A+\frac{1}{2}\sigma_M\right)} \tag{6}$$

Comparing Equation 3 and Equation 5, the cutoff for B is higher for the separating equilibrium relative to the partial separating equilibrium. This obtains because if B-informed traders were to trade the ETF, they would have to pay the adverse selection costs from A-informed traders.

Suppose, for example, that B-informed traders know the true value of security B is 1. They value A at the market maker's prior of δ, and therefore value the ETF at $\phi \cdot \delta + (1-\phi) \cdot 1$. In the partial separating equilibrium, the A-informed traders are mixing between A and the ETF. The market maker, anticipating this adverse selection from A-informed traders, sets the ETF ask at:

$$\phi\delta\frac{\mu_A\psi_1+\frac{1}{2}\sigma_M}{\delta\mu_A\psi_1+\frac{1}{2}\sigma_M}+(1-\phi)\beta$$

For B-informed investors, the trade-off between trading the ETF and trading stock B becomes:

$$\phi\left(\delta-\delta\frac{\mu_A\psi_1+\frac{1}{2}\sigma_M}{\delta\mu_A\psi_1+\frac{1}{2}\sigma_M}\right)+(1-\phi)(1-\beta)\leq 1-\beta\frac{\mu_B+\frac{1}{2}\sigma_B}{\beta\mu_B+\frac{1}{2}\sigma_B}$$

Note that $\phi\left(\delta-\delta\frac{\mu_A\psi_1+\frac{1}{2}\sigma_M}{\delta\mu_A\psi_1+\frac{1}{2}\sigma_M}\right)<0$, and this value represents the adverse selection that B-informed investors would have to pay when they trade the ETF. This adverse selection decreases the profitability of trading the ETF relative to trading stock B, leading to the higher cutoff values for mixing in Equation 5.

COROLLARY 2: *If A-informed investors mix between A and the ETF, investors with information about security B lose money by trading the ETF based on their knowledge of B so long as:*

Bid (B=0 and B traders consider selling the ETF): $$\phi\left(\frac{(1-\delta)\delta\mu_A\psi_2}{(1-\delta)\mu_A\psi_2+\frac{1}{2}\sigma_M}\right)\geq(1-\phi)\beta$$

Ask (B=1 and B traders consider buying the ETF): $$\phi\left(\frac{(1-\delta)\delta\mu_A\psi_1}{\delta\mu_A\psi_1+\frac{1}{2}\sigma_M}\right)\geq(1-\phi)(1-\beta)$$

The adverse selection from A-informed can become so severe that B-informed are completely excluded from the ETF. If the ETF were the only asset B-informed could trade, they would not make any trades. Their exclusion from the ETF occurs because investors in A have information that is more important to the ETF price. This importance of A information can come in two ways. First, A can have a larger ETF weight than B. Second, the potential change in value in A can be larger than the potential change in B. Together, both the weight and the volatility of A lead to a wide ETF spread on account of the A-informed. When B-informed traders value the ETF at a point between the bid and ask, they are excluded from trading the ETF.

When a trader with stock-specific information considers trading the ETF, they must consider both the availability of noise traders and the presence of traders with orthogonal information. When multiple assets are correlated with their information, they may not trade all assets. In Section IV A, informed traders with A-specific information only have to consider whether the ETF noise traders justify trading an asset which is less correlated with their information about stock A. When there are multiple pieces of information, traders must also consider whether the value of their information exceeds adverse selection from other traders. Even when a signal predicts an asset return better than the information contained in market prices, trading on the signal may not be profitable in the face of adverse selection from traders with other pieces of information.

This resultant exclusion underscores the asymmetric effect of ETFs on the underlying securities. Investors whose information is substantial—i.e. either about a large stock or predictive of a large value change—can profitably trade the ETF. But their presence creates adverse selection; in the ETF, this adverse selection screens out traders with information which is about small stocks or small value changes.

B. Fully Pooling Equilibrium

PROPOSITION 5: *If the conditions of Proposition 4 are violated for both securities, then there is a fully pooling equilibrium. Both traders trade the ETF, and following an ETF trade, the market maker has the following Bayesian posteriors:*

$$\delta_{buy} = \delta \frac{\mu_A \psi_{A,1} + \beta \mu_B \psi_{B,1} + \frac{1}{2}\sigma_M}{\delta \mu_A \psi_{A,1} + \beta \mu_B \psi_{B,1} + \frac{1}{2}\sigma_M} \qquad \delta_{sell} = \delta \frac{\frac{1}{2}\sigma_M}{(1-\delta)\mu_A \psi_{A,2} + (1-\beta)\mu_B \psi_{B,2} + \frac{1}{2}\sigma_M}$$

$$\beta_{buy} = \beta \frac{\delta \mu_A \psi_{A,1} + \mu_B \psi_{B,1} + \frac{1}{2}\sigma_M}{\delta \mu_A \psi_{A,1} + \beta \mu_B \psi_{B,1} + \frac{1}{2}\sigma_M} \qquad \beta_{sell} = \beta \frac{\frac{1}{2}\sigma_M}{(1-\delta)\mu_A \psi_{A,2} + (1-\beta)\mu_B \psi_{B,2} + \frac{1}{2}\sigma_M}$$

In a fully pooling equilibrium, both informed traders use a mixed strategy and trade both the stock and ETF. Following an ETF trade, the market maker updates beliefs about the value of both A and B.

In the fully pooling equilibrium, changes to stock-specific liquidity affect both securities. Suppose, for instance, that there is a reduction in σ_B the number of noise traders in Stock B, while all other portions of traders remain the same. Spreads in Stock B increase according as the ratio of informed to uninformed traders in Stock B increases. To reach a new equilibrium, B-informed traders send a higher portion of their trades to the ETF. This increases adverse selection in the ETF, and in response A-informed traders send a higher portion of their trades to Stock A, increasing spreads in Stock A. In the absence of the ETF, changes in the number of noise traders in Stock B has no effect on spreads in Stock A. With the ETF, informed traders in each stock access the same pool of liquidity in the ETF, and stock-specific changes have spillovers as a result.

The change in beliefs about each stock depends on more than just the stock weight in the ETF; the informational asymmetry, the portion of informed traders, and the market maker's uncertainty in belief about the stock all shape the updating process. Two stocks with equal weight in the ETF can see dramatically different adjustments in response to an ETF order. For example, if Stock A has more informed traders (μ_A high) and more uncertainty about fundamental value ($\delta \approx 0.5$) while Stock B has few informed traders (μ_B low and little uncertainty about fundamental value ($|\beta - 1| \approx 0$), following an ETF trade, the market maker's updated expected value about Stock A will change by more than the value of Stock B. Appendix D explores these properties in detail.

Each stock-specific property—the ETF weight, the population of traders, and the prior estimate of value—contributes to the market maker's updating process. The law of one price dictates that ETF and the sum of the underlying stocks should be equal, but it imposes no rigid rules about co-movement between the ETF and an individual stock. Even in a fully pooling equilibrium, ETF trades contribute to stock-specific price discovery, with some stocks seeing more adjustment than others in response to an ETF trade.

C. *Theoretical Predictions*

ETF trades contribute to the price discovery of stock-specific information, provided the stock's ETF weight or informational asymmetry is sufficiently large. In my model, the position limit given to informed traders allows the ETF to be non-redundant. For an A-informed investor, buying Stock A directly gives more exposure to stock A than buying the ETF for a given amount of capital or risk exposure. In a frictionless model, for all equilibria the ETF has a lower bid-ask spread than the underlying stocks, and offers a cheaper way for uninformed investors with market shocks to trade the market portfolio.[16] This lower ETF spread can attract Stock A-informed investors, but only when the ETF weight of Stock A is sufficiently large to justify committing capital or taking on factor risk by trading the ETF. In the multi-asset setting, traders also face adverse selection in the ETF. For example, if B-informed investors trade the ETF, the informational advantage of A-informed investors may be smaller than the ETF spread, in which case they do not trade the ETF regardless of their capital limit.

The difference in how large and small ETF-weight stocks interact with the ETF is no difference of degrees: informed investors in large-weight stocks trade the ETF while those in small-weight stocks do not. For small stocks, investors do not access ETF liquidity, and ETF trades are consequently uninformative about the idiosyncratic information of small stocks. For large stocks, however, ETF trades are informative about stock-specific information. Informed investors in large stocks can earn more profit trading both the stock and ETF than trading the stock alone. Effects are stronger

[16] Extending the model to include traders informed about the a common factor causes the ETF spread to potentially be equal to, but never greater than, the underlying stock spreads without additional non-informational frictions. This is because the market portfolio can always be replicated by trading only the ETF or trading only underlying stocks, while any arbitrary portfolio cannot be replicated by trading only the ETF. No traders, whether informed or noise traders, would pay more to trade at the ETF quote if they can obtain the same portfolio more cheaply by trading the underlying securities.

for stocks with larger ETF weights or larger informational asymmetries. For these stocks, ETF trades contribute to the price discovery of stock-specific information, and the harm of the ETF to stock-specific price discovery is limited.

VI. Empirics

To test the prediction that investors trade stocks and ETFs in tandem, I utilize microsecond timestamps from NYSE Trade and Quote (TAQ) data. While trades in the dataset are anonymous, I identify simultaneous trades, where a trade in a stock and a trade in an ETF occur within microseconds of each other. These trades occur too close in time to be responding to each other, and are far more numerous than would occur by chance (Figure 4).

I document that these simultaneous trades are common, comprising 1 to 2% of ETF volume. Market makers appear to view these simultaneous trades as well informed, as they have higher-than-average price impact. Simultaneous trades are driven by stock-specific information, as measured by intraday returns, earnings dates, and news article dates. Consistent with theoretical predictions, the effects are much stronger for larger stocks or larger informational asymmetries.

The sample of stocks and ETFs is described in Section A. In Section B, I define and measure simultaneous trades, and show they are driven by stock-specific information. These trades are of the same sign: investors buy the stock and ETF, or sell the stock and ETF, as analyzed in Section C. The volume and characteristics of simultaneous trades is reported in Section D. Remaining sections investigate news articles as a measure of information (Section E) and stock-ETF-stock triple trades (Section F). Further robustness checks are reported in the appendix and online appendix.

A. Data

Under my model, investors with stock-specific information trade ETFs whenever their stock has a sufficiently large ETF weight or has a sufficiently large informational asymmetry. To empirically test this, I examine links between stocks and ETFs, and I investigate how these links vary with stock-specific characteristics. For a sample of stocks, I take all stocks which appear in the S&P 500 between August 1, 2015 and December 31, 2018. For a sample of ETFs, I take all US-listed ETFs which have at least 50% of their holdings in individual stocks of the sample and trade on at least

10 days, yielding a sample of 552 ETFs.

Microsecond Trade and Quote (TAQ) data was collected for all stocks and ETFs in the sample. The sample period is from August 1, 2015 to December 31, 2018. These trades were cleaned according to Holden and Jacobsen (2014). ETF holdings were collected directly from State Street as well as from Master Data. Daily return data was obtained from CRSP. Intraday news events were collected from Ravenpack. Summary statistics on each stock are presented in Table II.

Table II: Summary Statistics on Securities

(a) Panel A: Stock Summary Statistics
My sample is comprised of all stocks listed in the S&P 500 Index between August 1, 2015 to December 21, 2018. During my sample period, there are 860 trading days.

Statistic	Mean	St. Dev.	5%	95%
Daily Simultaneous Trades	203	1,753	0	680
Daily ETF Orders	21,823	24,084	4,302	62,339
Daily Stock Orders	4,960	35,300	2	11,186
Intraday Stock Return	0.023	1.80	-2.67	2.53
Intraday ETF Return	0.023	0.89	-1.50	1.36

(b) Panel B: ETF Summary Statistics
Stock weights in an ETF are unequally distributed. The typical ETF has a few very large, concentrated holdings, and a numerous quantity of very small holdings. Across the 552 ETFs in my sample, an average of 45% of the ETF holdings are invested in assets with at least a 2% share in each asset.

Statistic	Mean	St. Dev.	5%	95%
Weight	0.5	1.1	0.02	2.3
Value Share of Stocks >5% Weight	14.1	21.5	0	61.0
Value Share of Stocks 2-5% Weight	30.7	29.2	0	98.4
Value Share of Stocks 1-2% Weight	19.6	20.4	0	61.1
Value Share of Stocks <1% Weight	35.5	35.0	0	100

B. Simultaneous Trades: Basic Setting

In this section, I test the model's prediction that investors with stock-specific information also trade ETFs. Under my theory, investors use this mixed strategy when both their single-stock informational asymmetry is high and their information has sufficient weight in the ETF. My identification of trade-both behavior relies on examining simultaneous trades. While TAQ data is anonymous, some traders who trade a stock and an ETF submit their orders at precisely the same time. As a result, I can identify their trading behavior. This idea is motivated by the measure of cross market activity proposed by Dobrev and Schaumburg (2017), which seeks to identify cross-market linkages through lead-lag relationships.

Rather than looking for a lead-lag between two markets, I look for simultaneous activity where trades occur too closely in time for one trade to be a response to another. Table III presents the latency between the gateway and limit order book for the major US exchanges. To respond to a trade, even the fastest co-located trading firms must pass through the gateway to the limit order book. All exchanges have a minimum latency of at least 20 microseconds; therefore, if two trades occur within 20 microseconds of each other, one trade cannot possibly be a response to the other. Using this physical limitation from the exchanges, I define trades as simultaneous if they occur within 20 microseconds of each other, and calculate the total number of such trades for each stock–ETF pairing[17].

To ensure that this raw measure of simultaneous trades is not influenced by increases in overall trading volume, I use the same baseline estimation correction suggested by Dobrev and Schaumburg (2017). This baseline estimate measures how many trades occur in both markets at a point close in time, but not exactly simultaneously. For each stock trade in my sample, I calculate how many ETF trades occur exactly X microseconds before or after the stock trade. I calculate the average number of trades as X ranges from 1000 to 1200 microseconds; this boundary is far enough away to avoid picking up high frequency trading response trades, but close enough to pick up patterns in trading at the millisecond level. As X in the simultaneous region ranges from $(-20, 20)$ while X in the baseline region ranges from $(-1200, -1000)$ and $(1000, 1200)$, I scale the baseline estimate by $\frac{20}{200}$. I then calculate simultaneous trades as the difference between the raw simultaneous measure and the scaled baseline measurement. The level of simultaneous trades between stock i and ETF j on day t can be written as:

$$\text{Simultaneous Trades}_{ijt} = \text{Raw Simultaneous}_{ijt} - \frac{20}{200}\text{Baseline}_{ijt}$$

[17] To ensure the trades take place within 20 microseconds of each other, I use the TAQ participant timestamp field and not SIP timestamps. All exchanges are required to timestamp their trades to the microsecond during my sample time period. Alternative trading facilities (ATF)'s, sometimes referred to as dark pools, are under less strict regulations, and only time stamp their trades to the nearest millisecond. As a result, I must exclude these millisecond-stamped trades from my analysis.

This baseline-corrected measure of simultaneous trades accounts for chance simultaneous trades which varies with changes in daily trading volume. Figure 4 plots a sample observation of cross market activity, along with the raw simultaneous and baseline regions.

Table III: **Gateway to Limit Order Book Latency from Major Exchanges**. This table gives the round-trip latencies reported by the major exchanges. All times are in microseconds, and reflect the total round-trip time from the gateway to limit order book of an exchange. All traders, including co-located high frequency traders, must make this trip. Note that IEX and NYSE American have significantly longer round-trip times due to the inclusion of a 350 microsecond "speed-bump."

	Minimum	Average
CBOE/BATS	31	56
Nasdaq	25	sub-40
NYSE	21	27
NYSE Arca	26	32
NYSE American	724	732
IEX	700+	700+

To test the model, the main regression examines how stock–ETF simultaneous trades change with stock-specific information. I consider two measures of stock-specific information: earnings dates and daily stock-specific return. This leads to two variations of the same regression:

REGRESSION 1: *For stock i, ETF j, and day t:*

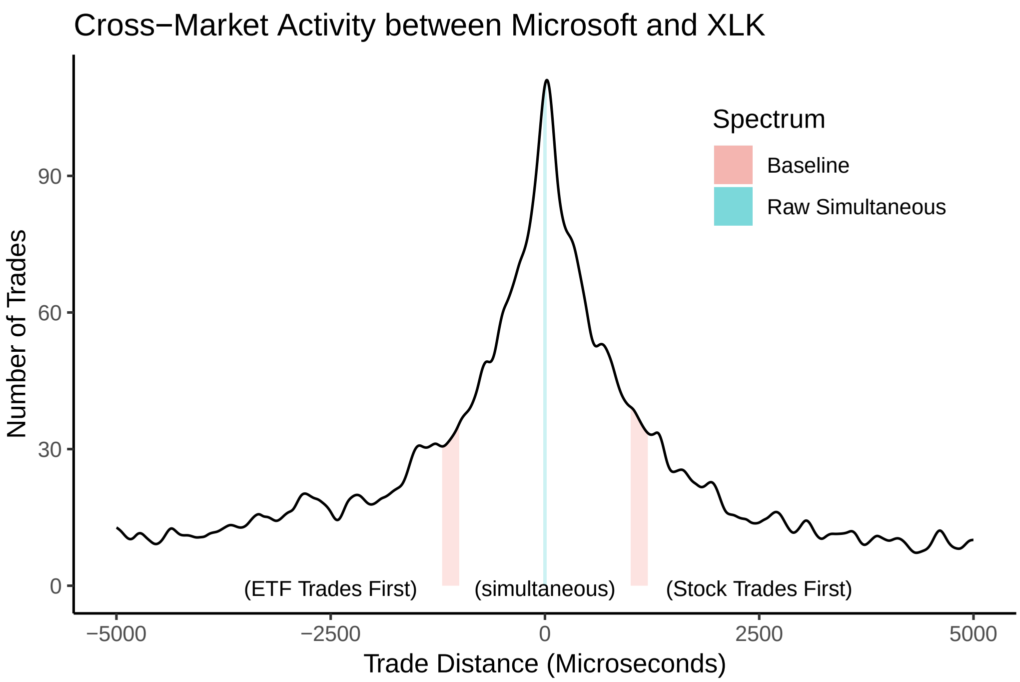
$$\begin{aligned} \textit{Simultaneous Trades}_{ijt} = \alpha_0 + \alpha_1 \textit{Earnings Date}_{it} + \alpha_2 \textit{Weight}_{ij} \\ + \alpha_3 \textit{Weight}_{ij} * \textit{Earnings Date}_{it} + \alpha_4 \textit{Controls}_{ijt} + \epsilon_{ijt} \end{aligned} \tag{7}$$

$$\begin{aligned} \textit{Simultaneous Trades}_{ijt} = \alpha_0 + \alpha_1 \textit{Abs Return}_{it} + \alpha_2 \textit{Weight}_{ij} \\ + \alpha_3 \textit{Weight}_{ij} * \textit{Abs Return}_{it} + \alpha_4 \textit{Controls}_{itj} + \epsilon_{ijt} \end{aligned} \tag{8}$$

Simultaneous Trades measures the number of simultaneous trades between stock i and ETF j on date t. Earnings Date is an indicator that takes the value of 1 on the trading day after a company releases earnings. Abs Return is the absolute value of the intraday return of a stock. Controls include a fixed effect for each ETF, the ETF return, and an interaction between stock weight and the ETF return.

Theory predicts a positive value for α_3. When a stock has a large weight in the ETF and there is a large amount of stock-specific information, investors should trade both the stock and the

Figure 4. Cross Market Activity in between Microsoft and XLK on October 2, 2018. The cross-market activity plot between Microsoft and the Technologies Sector SPDR XLK shows a clear spike in trades which occur at the same time. The x-axis depicts the offset of X microseconds between an XLK trade timestamp and Microsoft trade timestamp. The y-axis depicts the number of trade pairs which occur with that exact offset.
Trades in the thin blue region are the raw measure of simultaneous trades: when a stock and ETF trade less than 20 microseconds apart, physical limits from the exchange mean these trades cannot be responding to each other. To account for daily changes in overall number of trades, I estimate a baseline level of cross-market activity with the region in red, where trades in the ETF occur 1000 to 1200 microseconds before or after trades in the stock.



ETF. A semi-pooling or fully pooling equilibrium takes place only in the stocks that are sufficiently heavily-weighted in the ETF or have a sufficiently large informational asymmetry. When pooling does occur, the probability of submitting an order to the ETF is increasing in both the weight and the size of the informational asymmetry (Propositions 4 and 5). Results for Regression 1 are presented in Table IV.

The estimate of α_3, the interaction between weight and the size of the information, is positive for each of the measures of information considered in Regression 1. The increase in simultaneous trades is also sizeable; as an example, consider Exxon-Mobile, which comprises 20% of the Energies SPDR. Around earnings announcements, Exxon would have an additional 770 trades between Exxon and the ETF. For an extra 1% absolute return, Exxon would see an additional 425 simultaneous trades.

Table IV: **Estimation of Regression 1**

This table reports estimates of Regression 1, which estimates the effect of changes in stock-specific information on simultaneous trades between stocks and ETFs which contain them. I consider two different measures of stock-specific information: earnings dates and absolute value of the return. Earnings Date is an indicator which takes the value 1 for stocks which announce earnings before the day's trading session. Abs Return is the absolute value of the intraday return, measured as a percentage. The sample is all stocks listed in the S&P 500 index at any point August 1, 2015 to December 31, 2018, and the 552 ETFs selected for having at least 50% of their holdings invested in said stocks and trading on at least ten days. The frequency of observations is daily. I include a fixed effect for each ETF and cluster standard errors by ETF.

	Dependent variable: Simultaneous Trades			
	(1)	(2)	(3)	(4)
Weight	89.508***	90.163**	118.636***	96.546***
	(29.559)	(37.099)	(32.843)	(37.261)
Earnings Date			52.373	59.803
			(34.470)	(37.509)
Abs ETF Return		150.546*		163.476*
		(77.797)		(84.856)
Abs Return	39.419*	21.165*		
	(21.882)	(11.385)		
Weight∗Abs Return	34.683***	21.355***		
	(10.014)	(5.772)		
Weight∗Earnings Date			43.195**	38.486**
			(17.535)	(16.857)
Weight∗Abs ETF Return		14.408		30.110
		(20.932)		(22.466)
Observations	34,068,205	34,058,432	34,068,858	34,059,085
R^2	0.289	0.292	0.287	0.291
Residual Std. Error	1,479	1,476	1,481	1,476
Standard Errors Clustered by ETF.		*p<0.1; **p<0.05; ***p<0.01		

These increases are substantial in magnitude. On a day with return of less than 1%, Exxon averages 800 simultaneous trades with the Energies ETF. Relative to this number, earnings dates see a 100% increase in simultaneous trades while days with a 1% return see a 50% increase in simultaneous trades for each percentage point increase in return. And these results hold after controlling for the ETF return.

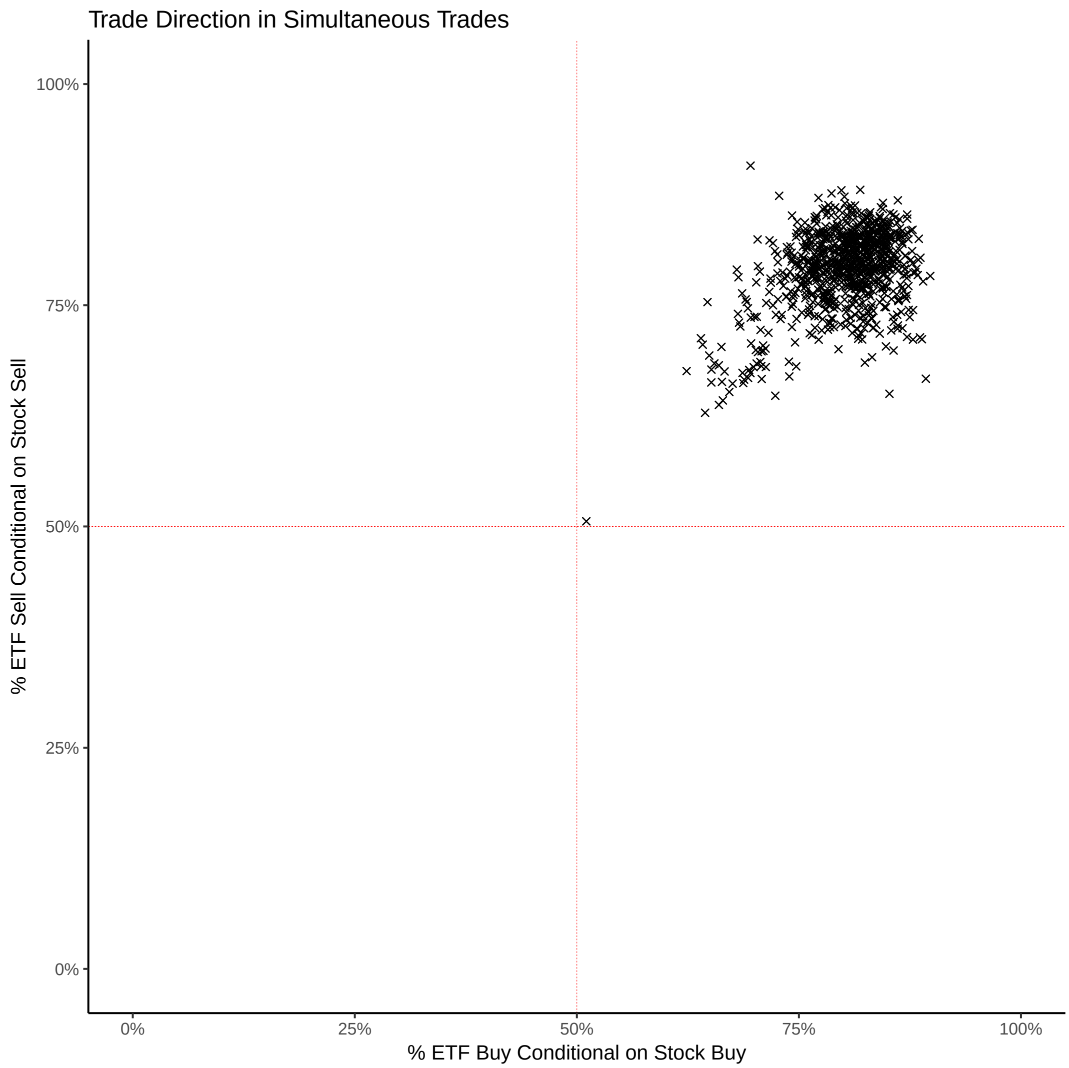
C. Signed Trades

One alternative explanation for the spike in cross market activity is that it merely represents hedging or "arbitrage," where an investor takes opposite positions in the two securities. Evidence from signed trades rules out this story. I show that these simultaneous trades are primarily simultaneous buy orders, where an investor buys both the stock and the ETF, or simultaneous sell orders, where an investor sells both the stock and the ETF.

I sign trades following Chakrabarty, Li, Nguyen, and Van Ness (2007), though results with trades signed according to Lee and Ready (1991) are similar. Figure 5 plots the percentage of simultaneous trades which have the same sign (ETF buys with stock buys or ETF sells with stock sells). With respect to both stock buy orders and stock sell orders, ETF trades overwhelmingly have the same sign. Trades in the same direction are inconsistent with arbitrage trades, which would require that investors buy one security while selling another. Attenuation from mis-signed trades would move data toward a 50-50 chance of buying or selling. Since almost all days are clustered around 80% of trades having the same sign, any inaccuracies in trade signing biases the results toward more mixed-sign trades, and fewer same-sign trades.

To confirm the intuition from Figure 5, I run a regression to analyze the effect of measures of stock-specific information on the change in simultaneous trades with the same direction. As before, substantial stock-specific information, either due to the size of the information asymmetry or the ETF weight, should lead to more simultaneous activity. In this specification, however, I use signed trades and check that there are more simultaneous trades of the same sign. For the stock sign $S_i = \{buy, sell\}$ and ETF sign $Y = \{buy, sell\}$, I run four regressions of the form:

Figure 5. Trade Direction. Simultaneous trades are overwhelmingly in the same direction: investors buy the stock and buy the ETF, or sell the stock and sell the ETF. Each point represents one day of an ETF-stock pairing. For the buy orders in the stock, the X-axis shows the percentage of simultaneous ETF trades which are also buy orders. For the sell orders in the stock, the X-axis shows the percentage of simultaneous ETF trades which are also sell orders. Thus the top right of the graph, at $(100, 100)$ would have all simultaneous trades from the day having the same sign, while the bottom left of the graph $(0, 0)$ would have all simultaneous trades having the opposite sign.



REGRESSION 2:

$$\begin{aligned}
Simultaneous\ Trades_S_Y_{ijt} &= \alpha_0 + \alpha_1 Earnings\ DateX_{ijt} + \alpha_2 Weight_{ijt} \\
&\quad + \alpha_3 Weight_{ijt} * Earnings\ DateX_{ijt} + \alpha_4 Controls_{ijt} + \epsilon_{ijt}
\end{aligned} \tag{9}$$

Results are presented in Table V. After an earnings announcement, stocks see a large increase in simultaneous trades in the same direction. Investors buy both the stock and the ETF at the same time, or sell both the stock and the ETF at the same time. The greater the weight of the stock in the ETF, the larger this increase in simultaneous trades of the same sign. I also estimate Regression V for the absolute stock return, and find similar results.

Finally, it is worth noting that these results do not rule out that hedging occurs, but only that confirms that the bulk of simultaneous trades are not hedging. If an investor is seeking to buy both a stock and an ETF which contains that stock, executing the trades simultaneously is

Table V : Estimation of Regression 2 - Signed Trades

This table reports estimates of Regression 2, which estimates the effect of stock-specific information on stock–ETF simultaneous trades of a particular trade sign. Column (1) estimates the effect for simultaneous buy orders. Column (2) estimates the effect for simultaneous trades comprised of a buy order in the stock with a sell order in the ETF. Column (3) estimates the effect for a sell order in the stock with a buy order in the ETF. Column (4) estimates simultaneous sell orders. Earnings Date is an indicator which takes the value 1 for stocks which announce earnings before the day's trading session. Abs Return is the absolute value of the intraday return, measured as a percentage. The sample is all stocks listed in the S&P 500 index at any point August 1, 2015 to December 31, 2018, and the 552 ETFs selected for having at least 50% of their holdings invested in said stocks and trading on at least ten days. The frequency of observations is daily. I include a fixed effect for each ETF and cluster standard errors by ETF.

Panel A: Regression Estimates with Absolute Return

	Dependent variable: Simultaneous Trades			
	(1)	(2)	(3)	(4)
Stock Trade Sign:	BUY	BUY	SELL	SELL
ETF Trade Sign:	BUY	SELL	BUY	SELL
Weight	29.035***	17.985**	18.727**	28.834**
	(11.056)	(8.312)	(8.501)	(11.457)
Abs Return	6.304*	4.903*	4.895*	6.293*
	(3.300)	(2.838)	(2.786)	(3.325)
Abs ETF Return	44.490**	30.398*	32.836*	50.504**
	(21.771)	(17.405)	(19.168)	(24.354)
Weight∗Abs Return	6.490***	4.436***	4.606***	6.854***
	(1.690)	(1.279)	(1.313)	(1.739)
Weight∗Abs ETF Return	5.116	3.023	2.790	5.164
	(6.280)	(4.606)	(5.016)	(6.610)
Observations	35,038,413	35,116,833	35,027,361	35,106,126
R^2	0.326	0.268	0.256	0.296
Residual Std. Error	409	331.720	349	440
Standard Errors Clustered by ETF.				*p<0.1; **p<0.05; ***p<0.01

Table V: (continued from previous page)

Panel B: Regression Estimates With Earnings Dates

	Dependent variable: Simultaneous Trades			
	(1)	(2)	(3)	(4)
Stock Trade Sign:	BUY	BUY	SELL	SELL
ETF Trade Sign:	BUY	SELL	BUY	SELL
Weight	31.030***	19.284**	20.091**	30.919***
	(11.117)	(8.301)	(8.546)	(11.563)
Earnings Date	17.770	16.990	16.345	16.174
	(11.053)	(10.877)	(10.136)	(10.707)
Abs ETF Return	48.349**	33.413*	35.837*	54.336**
	(23.826)	(19.180)	(20.904)	(26.410)
Weight∗Earnings Date	9.445**	7.749*	7.513**	13.339***
	(4.387)	(4.132)	(3.670)	(5.057)
Weight∗Abs ETF Return	9.926	6.318	6.209	10.232
	(6.796)	(4.984)	(5.322)	(6.985)
Observations	35,039,078	35,117,496	35,028,028	35,106,791
R^2	0.326	0.268	0.256	0.295
Residual Std. Error	409.342	331.844	349.869	440.777

Standard Errors Clustered by ETF. *p<0.1; **p<0.05; ***p<0.01

important, as trading in one asset could push up the cost of the other. If investors are using ETFs to hedge industry exposure, as in Huang et al. (2021), there is no need to execute the trades within microseconds of each other. If anything, an investor would want to wait to take the hedging position, as a buy order in the stock may push up the ETF price prior to their selling the ETF (and vice versa).

D. Effect of Simultaneous Trades

The trade characteristics of simultaneous trades are consistent with these trades being informationally-motivated. I consider two measures: price impact and realized spread. Price impact measures the change in price per unit of volume. Realized spread is the difference between the trade price and the price five minutes after the trade. Table VIII summarizes these trade characteristics. Simultaneous trades have a much larger price impact in both the ETF and the single stock. Realized spreads for simultaneous trades are negative, suggesting the traders who place these orders earn immediate trading profits of around .75 cents per share on the ETF side of the simultaneous trade, and 1.4 cents per share on the stock side of the trade. This is consistent with the idea that these simultaneous trades are placed by informed traders who increase their profits by trading both a stock and the ETF.

The total size of these simultaneous trades is also considerable. Table VII presents volume shares for Sector SPDRs, while Table VI presents the volume shares for SPY. As much as 1% to 2% of sector ETF volume comes from these simultaneous orders from a single stock. In SPY, simultaneous trades with a stock like Apple or Microsoft can comprise 0.3% to 0.8% of daily ETF volume. All these numbers are obtained by taking the raw simultaneous volume and subtracting off a baseline level of cross-market activity estimated from trades that take place between 1 and 2 microseconds apart.

Table VI: SPY Volume Shares. This table shows the volume share of simultaneous trading activity in each ETF for stocks which have at least 1.5% weight in SPY. For each stock, I give the 10^{th}, 50^{th} and 90^{th} percentiles of daily observations of this share. For example, on a median day 0.33% of the daily trading volume in SPY occurs in trades that are simultaneous with trades in Apple.

Stock	ETF	10%	50%	90%
AAPL	SPY	0.19	0.33	0.68
MSFT	SPY	0.15	0.25	0.86
XOM	SPY	0.05	0.1	0.15
JNJ	SPY	0.06	0.1	0.18
GE	SPY	0.05	0.08	0.15
FB	SPY	0.11	0.2	0.4
AMZN	SPY	0.05	0.08	0.19
JPM	SPY	0.12	0.26	0.46
GOOG	SPY	0.06	0.1	0.13

Table VII: Sector SPDR Volume Shares. This table shows the volume share of simultaneous trading activity in each ETF for stocks which have at least 10% weight in a Sector SPDR. For each stock, I give the 10^{th}, 50^{th} and 90^{th} percentiles of daily observations of this share. For example, on a median day, 1.15% of the daily trading volume in XLK occurs in trades that are simultaneous with trades in Apple.

Stock	ETF	10%	50%	90%
AAPL	XLK	0.74	1.15	1.92
MSFT	XLK	0.69	1.19	2.49
CVX	XLE	0.30	0.44	0.66
XOM	XLE	0.30	0.48	0.83
GE	XLI	0.30	0.45	0.63
PG	XLP	0.34	0.52	0.75
DD	XLB	0.24	0.33	0.49
DOW	XLB	0.32	0.47	0.67
DWDP	XLB	0.55	0.8	1.08
LIN	XLB	0.16	0.23	0.25
JNJ	XLV	0.29	0.45	0.71
AMZN	XLY	0.38	0.7	0.97
HD	XLY	0.28	0.37	0.48
JPM	XLF	0.55	0.92	1.46
WFC	XLF	0.53	0.53	0.53
PM	XLP	0.38	0.38	0.38
KO	XLP	0.38	0.51	0.73
FB	XLC	1.52	2.12	3.10
GOOG	XLC	0.75	0.95	1.25
NEE	XLU	0.31	0.54	0.77

Table VIII Summary Of Trade Characteristics. This table reports mean price impacts and mean realized spreads for trades, measured in cents. Simultaneous trades have much higher price impacts and earn negative realized spreads, consistent with market makers viewing these trades as well informed.

	All Trades	Simultaneous Trades
ETF Price Impact	$-.541$	3.12
Stock Price Impact	2.91	4.03
ETF Realized	8.08	-.756
Stock Realized	.087	-1.407

E. *Intraday News Events*

As an alternative measure of stock-specific news events, I use the Dow Jones Edition and PR Edition data from Ravenpack Analytics. These datasets aggregate news wires and media articles about companies. As Ravenpack does not collect articles for all ETFs, I restrict analysis to SPDR and the Sector SPDR ETFs. From August 2015 to December 2018, there are 136,000 unique news articles related to the stocks in my sample. Ravenpack data allows analysis of intra-day trading patterns. For a stock which has a news event, I calculate—over 5 and 30 minute intervals both before and after the news event—simultaneous activity between that stock and the Sector SPDR. I also calculate these measures of stock-Sector SPDR simultaneous trades for each of the other stocks from the Sector SPDR of the stock which had the news event. This allows a comparison between the stock which has the news event against the other stocks in the sector.

I re-estimate Regression 4 for the 5 and 30 minute intervals before and after each news event. Information is now an indicator which takes the value of one for the stock in the sector SPDR which had the news event. I also include a fixed effect for each news event.

Table IX: **Estimation of Regression 4 Using Ravenpack News Events**
This table reports estimates of Regression 4 around news events. For each company-specific news event, I take all stocks from that company's GICS sector, and measure all the stock-Sector SPDR simultaneous trades over 5 and 30 minute intervals both before and after the event. *Stock News* is an indicator which takes the value of 1 for the company which had the news event. Small stocks are stocks with less than 2% ETF weight in a Sector SPDR. Medium stocks have 2 to 5% ETF weight, and large stocks have an ETF weight greater than 5%. Controls include a fixed effect for each ETF, a fixed effect for each stock category, as well as a fixed effect for each news event. Standard Errors are clustered by ETF.

	Total Trades Per Hour			
	Before		*After*	
	30 Min	5 Min	5 Min	30 Min
Small Stock News	−0.056	0.267	−0.029	-2.615^{***}
	(0.150)	(0.284)	(0.376)	(0.366)
Medium Stock News	0.291^{**}	1.349^{***}	1.462^{***}	1.547^{**}
	(0.131)	(0.505)	(0.442)	(0.612)
Large Stock News	3.350^{**}	9.206	9.062^{*}	17.074^{***}
	(1.649)	(5.707)	(4.737)	(4.679)
Observations	8,825,809	8,825,809	8,825,809	8,825,809
R^2	0.446	0.299	0.304	0.379
Residual Std. Error	8.874	47.356	48.421	11.607
Standard Errors Clustered by ETF:			*p<0.1; **p<0.05; ***p<0.01	

Results are presented in Table IX. When small stocks have a news event, there is no increase in simultaneous trading at any time horizon. This is consistent with the prediction that traders with small-stock information face exclusion from the ETF due to the adverse selection they face from traders with more substantial information. For medium stocks, there is a modest increase in simultaneous trades around the event, with an extra 1 to 2 simultaneous trades per hour. Large stocks have a substantial increase in simultaneous activity over a wide time horizon. When a large stock has a news story, there are an additional 17 simultaneous trades per hour relative to the other stocks in that sector. Large stocks typically see 30 to 60 simultaneous trades per hour, so this represents a 25 to 50% increase over the normal rate of simultaneous trades.

F. Stock-ETF-Stock Triples

I consider triple-trades between a stock, an ETF, and a secondary stock within that ETF. These trades allow analysis of the importance of correlated information in determining simultaneous stock-ETF trades. In the model, investors have information which is purely idiosyncratic. In reality, no information meets this criteria: even information which appears to be specific to one company is relevant to that company's competitors. To focus on cases of this correlated information, I use the ten sector SPDR ETFs, where each ETF holds stocks from a given sector. If payoffs are positively correlated, investors have an additional incentive to trade the ETF, and may have an incentive to trade a secondary stock within that ETF.

To investigate this behavior, I develop the measure $\textit{Simultaneous Triple}_{ijkt}$. For all trades in stock i on day t, I measure how many are within 20 microseconds of both a trade in the Sector SPDR j and a secondary stock, k from that sector. I compute this measure for each firm pairing i and k from each sector j, and run the following regression:

REGRESSION 3: *For stock i, ETF j, secondary stock k, and day t:*

$$\begin{aligned} \textit{Simultaneous Trade}_{ijt} = \alpha_0 &+ \alpha_1 \textit{Abs Return}_{it} + \alpha_2 \textit{Weight}_{ij} + \alpha_3 \textit{Secondary Weight}_{kj} \\ &+ \alpha_4 \textit{Weight}_{ij} * \textit{Abs Return}_{it} + \alpha_5 \textit{Secondary Abs Return}_{kt} \\ &+ \alpha_6 \textit{Secondary Weight}_{kj} * \textit{Secondary Abs Return}_{kt} \\ &+ \alpha_7 \textit{Controls}_{itj} + \epsilon_{ijkt} \end{aligned} \tag{10}$$

$$\begin{aligned} \textit{Simultaneous Triple}_{ijkt} = \alpha_0 &+ \alpha_1 \textit{Abs Return}_{it} + \alpha_2 \textit{Weight}_{ij} + \alpha_3 \textit{Secondary Weight}_{kj} \\ &+ \alpha_4 \textit{Weight}_{ij} * \textit{Abs Return}_{it} + \alpha_5 \textit{Secondary Abs Return}_{kt} \\ &+ \alpha_6 \textit{Secondary Weight}_{kj} * \textit{Secondary Abs Return}_{kt} \\ &+ \alpha_7 \textit{Controls}_{itj} + \epsilon_{ijkt} \end{aligned} \tag{11}$$

Secondary Weight is the ETF weight of stock k in Sector SPDR ETF j. Other key variables are defined as in Regression 6. In Regression 3, I estimate two variations. Equation 10 uses simultaneous trades between stock i and ETF j as the dependent variable, while Equation 11 uses the triple of a simultaneous trade between stock i, ETF j, and secondary stock k from the same sector. Results are presented in Table X.

For the stock i and ETF j pairing, the interaction α_5 between the stock weight and absolute return is strongly predictive of the simultaneous double trade, while the interaction α_6 between the secondary stock weight and secondary absolute return is only weakly predictive relative to α_5. In contrast, for the triple pairing of stock i, ETF j, and secondary stock k, the interaction terms α_5 and α_6 are comparable in magnitude. In fact, α_6 is even larger than α_5, and thus the interaction between the secondary stock return and the absolute value of the secondary stock return is a key driver of the simultaneous stock-Sector SPDR-secondary stock triple.

As a practical example, consider two stocks, Apple and Microsoft, from the Technology Sector SPDR ETF (XLK). In predicting simultaneous Apple-XLK trades, small changes in the return of Apple have a large impact on Apple-XLK simultaneous trades, while changes in the return of Microsoft have a very small impact on Apple-XLK simultaneous trades. For simultaneous Apple-XLK-Microsoft trades, however, the return on Microsoft has a similar impact on the number of simultaneous triple-trades as that of the return on Apple, suggesting investors only trade all three when they have information pertinent to both Apple and Microsoft.

Table X: **Estimation of Regression 3**

This table reports estimates of Regression 3, which estimates the effect of changes in stock-specific information on simultaneous trades between either the double of a stock-Sector SPDR ETF pairing, or the triple between a stock, Sector SPDR ETF, and secondary stock from that index. Column (1) reports estimates for the double pairing, while Column (2) reports estimates for the triple. Weight is the stock weight in its Sector SPDR ETF, while Secondary weight is the weight of the secondary stock in the ETF. Abs Return is the absolute value of the intraday return, measured as a percentage. I control for the ETF return with the ETF return net of the return of the primary and secondary stocks. The sample is SPDR and the ten Sector SPDR ETFs and their stock constituents from August 1, 2015 to December 31, 2018. The frequency of observations is daily. I include a fixed effect for each ETF and cluster standard errors by ETF.

	Dependent variable:	
	Simultaneous Double	Simultaneous Triple
	(1)	(2)
Weight	23.866**	0.588***
	(9.993)	(0.205)
Abs ETF Return	23.924***	0.575**
	(2.796)	(0.277)
Abs Return	0.994	0.094
	(2.314)	(0.213)
Secondary Weight	0.078	0.564***
	(0.189)	(0.115)
Abs Secondary Return	1.873	−0.061
	(2.622)	(0.264)
Weight∗Abs ETF Return	3.993	0.154***
	(3.307)	(0.055)
Weight∗Abs Return	9.039***	0.215***
	(2.028)	(0.021)
Secondary Weight ∗Abs Secondary Return	0.548***	0.357***
	(0.171)	(0.108)
Observations	3,685,179	3,686,559
R^2	0.373	0.273
Adjusted R^2	0.373	0.273
Residual Std. Error	138.169 (df = 3685168)	7.814 (df = 3686548)
Note:		*p<0.1; **p<0.05; ***p<0.01

VII. Conclusion

My paper theoretically models and empirically demonstrates when investors strategically trade stocks and ETFs in tandem, and how this strategic behavior leads to stock-specific price discovery from ETFs. Trading stocks and ETFs in tandem allows investors with stock-specific information to reduce the market impact of their trades and increase profits. This trade-both behavior attenuates concerns about the impact of ETFs on price discovery. If noise traders do move to ETFs, informed traders follow them. Profitable trading opportunities, as well as the requisite acquisition of private information, are maintained. Rigid co-movement is avoided; following an ETF trade, market makers have flexibility in updating quotes. Stocks with a more certain value or a low level of informed trading receive small quote changes, while stocks with an uncertain value or a high level of informed trading receive larger quote changes.

The stock–ETF relationship varies depending upon both the stock weight in the ETF and the level of asymmetric information in the stock. Viability of the trade-both strategy requires both a sufficiently high weight of the stock in the ETF and a large enough information asymmetry. When these conditions are met, investors trade both the stock and the ETF—and they trade both in the same direction. This behavior creates adverse selection in the ETF, whereby different pieces of information in the ETF behave as substitutes. On the other hand, traders with information about small stocks or small information asymmetries can be excluded from the ETF whenever their information's value in pricing the ETF is less than the cost of adverse selection from other traders in the ETF.

The ETF weight of a stock will vary across ETFs, but the relative effects remain. Consider two companies, Amazon and American Tower Corporation. Being a large stock in a large sector, Amazon comprises a substantial 22% of the Consumer Discretionary Sector ETF, and a modest 4% of the SPDR ETF. In contrast, American Tower Corporation, a large stock in a small sector, comprises 14% of the Real Estate Sector ETF but less than 0.4% of SPDR. Investors with large informational advantage in Amazon find themselves with sizable informational advantages and profitable trading opportunities in both ETFs, while investors with an informational advantage in American Tower Corporation have little informational advantage in SPDR.

With precise timestamp data, I am able to exploit exchange latencies to identify investors who trade both the stock and the ETF simultaneously. These simultaneous trades have the same sign—buys in both markets or sells in both markets—which is inconsistent with an arbitrage story. Simultaneous trades have larger price impacts than average trades, and earn negative realized spreads on both the stock and ETF portion of the trade. Market makers appear to view these simultaneous trades as well informed.

Large stock-specific information—as measured by earnings dates, returns, or news events—leads to large increases in single-stock–ETF simultaneous trades. This increase in simultaneous trades does not show up between a stock and an unrelated ETF, and does not appear to be sensitive to the distribution of trades across time. As the model predicts, effects are stronger both in larger stocks and larger informational events, as measured by the realized return. The overall volume is significant, with the ETF side of trades simultaneous with a large single name frequently comprising 1% to 2% of ETF daily volume.

The price discovery process takes place across multiple assets and many exchanges. I show that ETFs are an important venue for stock-specific price discovery. Potential harms from ETFs rely on the assumption that ETFs completely screen out stock-specific information. I show that these harms are mainly localized to small stocks and small informational asymmetries. When investors have stock-specific information which is substantial on account of the ETF weight or the size of the informational asymmetry, investors with this stock-specific information trade both stocks and ETFs. In these settings, I conclude that ETFs can provide stock-specific price discovery.

REFERENCES

Admati, Anat R, and Paul Pfleiderer, 1988, Selling and trading on information in financial markets, *The American Economic Review* 78, 96–103.

Aquilina, Matteo, Eric Budish, and Peter O'Neill, 2020, Quantifying the high-frequency trading "arms race": A simple new methodology and estimates, *United Kingdom Financial Conduct Authority Occasional Paper* .

Back, Kerry, 1993, Asymmetric information and options, *The Review of Financial Studies* 6, 435–472.

Back, Kerry, C Henry Cao, and Gregory A Willard, 2000, Imperfect competition among informed traders, *The Journal of Finance* 55, 2117–2155.

Bae, Kyounghun, and Daejin Kim, 2020, Liquidity risk and exchange-traded fund returns, variances, and tracking errors, *Journal of Financial Economics* 138, 222–253.

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

Bessembinder, Hendrik, Chester Spatt, and Kumar Venkataraman, 2019, A survey of the microstructure of fixed-income markets, *Journal of Financial and Quantitative Analysis* 1–90.

Bhattacharya, Ayan, and Maureen O'Hara, 2016, Can etfs increase market fragility? effect of information linkages in etf markets .

Bond, Philip, and Diego Garcia, 2018, The equilibrium consequences of indexing, Technical report, Working paper.

Box, Travis, Ryan Davis, Richard Evans, and Andrew Lynch, 2021, Intraday arbitrage between etfs and their underlying portfolios, *Journal of Financial Economics* .

Brogaard, Jonathan, Davidson Heath, and Da Huang, 2019, Etf trading and the bifurcation of liquidity, *Available at SSRN 3510359* .

Cespa, Giovanni, and Thierry Foucault, 2014, Illiquidity contagion and liquidity crashes, *The Review of Financial Studies* 27, 1615–1660.

Chakrabarty, Bidisha, Bingguang Li, Vanthuan Nguyen, and Robert A Van Ness, 2007, Trade classification algorithms for electronic communications network trades, *Journal of Banking & Finance* 31, 3806–3821.

Chinco, Alex, and Vyacheslav Fos, 2021, The sound of many funds rebalancing, *The Review of Asset Pricing Studies* 11, 502–551.

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

Dannhauser, Caitlin D, 2017, The impact of innovation: Evidence from corporate bond exchange-traded funds (etfs), *Journal of Financial Economics* 125, 537–560.

Dannhauser, Caitlin D, and Saeid Hoseinzade, 2022, The unintended consequences of corporate bond etfs: Evidence from the taper tantrum, *The Review of Financial Studies* 35, 51–90.

Dobrev, Dobrislav, and Ernst Schaumburg, 2017, High-frequency cross-market trading: Model free measurement and applications, *Perspectives* .

Donohue, Dorothy, 2012, Re: Nasdaq stock market llc proposal with respect to inav pegged orders for etfs (file no. sr- nasdaq-2012-117).

Easley, David, David Michayluk, Maureen O'Hara, and Tālis J Putniņš, 2021, The active world of passive investing, *Review of Finance* 25, 1433–1471.

Easley, David, and Maureen O'Hara, 1987, Price, trade size, and information in securities markets, *Journal of Financial economics* 19, 69–90.

Evans, Richard B, Rabih Moussawi, Michael S Pagano, and John Sedunov, 2019, Etf short interest and failures-to-deliver: Naked short-selling or operational shorting? .

Foster, F Douglas, and S Viswanathan, 1996, Strategic trading when agents forecast the forecasts of others, *The Journal of Finance* 51, 1437–1478.

Glosten, Lawrence R, and Paul R Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71–100.

Glosten, Lawrence R, Suresh Nallareddy, and Yuan Zou, 2016, Etf activity and informational efficiency of underlying securities, *Columbia Business School Research Paper* .

Goldstein, Itay, Yan Li, and Liyan Yang, 2013, Speculation and hedging in segmented markets, *The Review of Financial Studies* 27, 881–922.

Gorton, Gary, and George Pennacchi, 1991, Security baskets and index-linked securities, Technical report, National Bureau of Economic Research.

Haddad, Valentin, Alan Moreira, and Tyler Muir, 2021, When selling becomes viral: Disruptions in debt markets in the covid-19 crisis and the fed's response, *The Review of Financial Studies* 34, 5309–5351.

Hamm, Sophia JW, 2014, The effect of etfs on stock liquidity .

Hasbrouck, Joel, 1995, One security, many markets: Determining the contributions to price discovery, *The journal of Finance* 50, 1175–1199.

Hasbrouck, Joel, 2003, Intraday price formation in us equity index markets, *The Journal of Finance* 58, 2375–2400.

Hasbrouck, Joel, 2018, Price discovery in high resolution, *Available at SSRN 3288754* .

Hasbrouck, Joel, and Gideon Saar, 2013, Low-latency trading, *Journal of Financial Markets* 16, 646–679.

Holden, Craig W, and Stacey Jacobsen, 2014, Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions, *The Journal of Finance* 69, 1747–1785.

Holden, Craig W, Yifei Mao, and Jayoung Nam, 2018, Price discovery in the stock, otc corporate bond, and nyse corporate bond markets, *Kelley School of Business Research Paper* .

Holden, Craig W, and Jayoung Nam, 2019, Market accessibility, corporate bond etfs, and liquidity .

Huang, Shiyang, Maureen O'Hara, and Zhuo Zhong, 2021, Innovation and informed trading: Evidence from industry etfs, *The Review of Financial Studies* 34, 1280–1316.

Israeli, Doron, Charles MC 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.

Johnson, Travis L, and Eric C So, 2012, The option to stock volume ratio and future returns, *Journal of Financial Economics* 106, 262–286.

Lee, Charles MC, and Mark J Ready, 1991, Inferring trade direction from intraday data, *The Journal of Finance* 46, 733–746.

Lu, Zhongjin, and Zhongling Qin, 2021, Leveraged funds and the shadow cost of leverage constraints, *The Journal of Finance* 76, 1295–1338.

Mackintosh, Phil, 2014, Etfs rarely enter the arb zone.

Madhavan, Ananth, and Aleksander Sobczyk, 2016, Price dynamics and liquidity of exchange-traded funds, *Journal of Investment Management* 14, 1–17.

Madhavan, Ananth, Aleksander Sobczyk, and Andrew Ang, 2019, What happens with more funds than stocks?, *Available at SSRN 3465888* .

Malamud, Semyon, 2016, A dynamic equilibrium model of etfs .

Pan, Kevin, and Yao Zeng, 2016, Arbitrage under liquidity mismatch: Theory and evidence from the authorized participants of bond etfs, Technical report, Working Paper.

Ross, Stephen A, 2013, The arbitrage theory of capital asset pricing, in *Handbook of the fundamentals of financial decision making: Part I*, 11–30 (World Scientific).

Sağlam, Mehmet, Tugkan Tuzun, and Russ Wermers, 2019, Do etfs increase liquidity?, *Available at SSRN 3142081* .

Subrahmanyam, Avanidhar, 1991, A theory of trading in stock index futures, *The Review of Financial Studies* 4, 17–51.

Yu, Lei, 2005, Basket securities, price formation, and informational efficiency, *Price Formation, and Informational Efficiency (March 25, 2005)* .

Appendix A. Proofs

Proof of Proposition 2: *In a pooling equilibrium, informed investors mix between A and the ETF and submit orders to the ETF with the following probability:*

$$\text{ETF Buy Probability (A=1): } \psi_1 = \frac{\phi\delta\mu_A\sigma_M - \frac{1}{2}(1-\phi)\sigma_A\sigma_M}{\delta\mu_A[\sigma_A + \phi\sigma_M]}$$
$$\text{ETF Sell Probability (A=0): } \psi_2 = \frac{\phi(1-\delta)\mu_A\sigma_M - \frac{1}{2}(1-\phi)\sigma_A\sigma_M}{(1-\delta)\mu_A[\sigma_A + \phi\sigma_M]}$$

For the informed trader to be willing to mix, he must be indifferent between buying the ETF or buying the individual stock. When $A = 1$, ψ_1 must solve:

$$\begin{aligned}
\phi + (1-\phi)\frac{1}{2} - ask_{(AB)} &= 1 - ask_A \\
\phi(1-\delta)\frac{\mu(1-\psi_1) + (1-\mu)\frac{1}{2}(1-\sigma_A)}{\delta\mu(1-\psi_1) + (1-\mu)\frac{1}{2}(1-\sigma_A)} &= 1 - \delta\frac{\mu\psi_1 + (1-\mu)\frac{1}{2}\sigma_A}{\delta\mu\psi_1 + (1-\mu)\frac{1}{2}\sigma_A} \\
\phi\frac{(1-\sigma_A)}{\delta\mu(1-\psi_1) + (1-\mu)\frac{1}{2}(1-\sigma_A)} &= \frac{\sigma_A}{\delta\mu\psi_A + (1-\mu)\frac{1}{2}\sigma_A} \\
\psi_1 &= \frac{\delta\mu\phi(1-\sigma_A) - \frac{1}{2}(1-\mu)(1-\sigma_A)\sigma_A(1-\phi)}{\mu\delta[\sigma_A + \phi(1-\sigma_A)]}
\end{aligned}$$

When $A = 0$, we have the following indifference condition for mixing:

$$\begin{aligned}
bid_{(AB)} - (1-\phi)\frac{1}{2} &= bid_A \\
\phi\delta\frac{(1-\mu)\frac{1}{2}(1-\sigma_A)}{(1-\delta)\mu(1-\psi_2) + (1-\mu)\frac{1}{2}(1-\sigma_A)} &= \delta\frac{(1-\mu)\frac{1}{2}\sigma_A}{(1-\delta)\mu\psi_2 + (1-\mu)\frac{1}{2}\sigma_A} \\
(1-\sigma_A)(1-\delta)\mu\psi_2 + (1-\sigma_A)(1-\mu)\frac{1}{2}\sigma_A &= (1-\delta)\mu(1-\psi_2)\sigma_A + (1-\mu)\frac{1}{2}(1-\sigma_A)\sigma_A \\
\psi_2 &= \frac{(1-\delta)\mu\phi(1-\sigma_A) - \frac{1}{2}(1-\mu)(1-\sigma_A)\sigma_A(1-\phi)}{\mu(1-\delta)[\sigma_A + \phi(1-\sigma_A)]}
\end{aligned}$$

Note that the condition for $\psi > 0$ is the same as the cutoff values for a pooling equilibrium to exist.

Proof of Proposition 3: *If there are ζ_A stock A noise traders in the absence of the ETF,* *then informed trader profits are lower in the presence of the ETF so long as:*

$$\zeta_A \leq \frac{\mu_A(1-\delta)[\sigma_A + \phi\sigma_M]}{\mu_A(1-\delta) - \frac{1}{2}(1-\phi)\sigma_M}$$

Comparing investor profits with and without the ETF:

$$\begin{aligned}
\delta\frac{\frac{1}{2}\varsigma_A}{(1-\delta)\mu_A + \frac{1}{2}\varsigma_A} &\leq \delta\frac{\frac{1}{2}\sigma_A}{(1-\delta)\mu_A(1-\psi_2) + \frac{1}{2}\sigma_A} \\
\varsigma_A\bigg((1-\delta)\mu_A(1-\psi_2) + \frac{1}{2}\sigma_A\bigg) &\leq \sigma_A\bigg((1-\delta)\mu_A + \frac{1}{2}\varsigma_A\bigg) \\
\varsigma_A(1-\delta)\mu_A(1-\psi_2) + \varsigma_A\frac{1}{2}\sigma_A &\leq \sigma_A(1-\delta)\mu_A + \sigma_A\frac{1}{2}\varsigma_A \\
\varsigma_A(1-\delta)\mu_A(1-\psi_2) &\leq \sigma_A(1-\delta)\mu_A \\
\varsigma_A(1-\psi_2) &\leq \sigma_A \\
\varsigma_A\left(1 - \frac{\phi(1-\delta)\mu_A\sigma_M - \frac{1}{2}(1-\phi)\sigma_A\sigma_M}{\mu_A(1-\delta)[\sigma_A + \phi\sigma_M]}\right) &\leq \sigma_A \\
\varsigma_A\frac{\mu_A(1-\delta)\sigma_A - \frac{1}{2}(1-\phi)\sigma_A\sigma_M}{\mu_A(1-\delta)[\sigma_A + \phi\sigma_M]} &\leq \sigma_A \\
\varsigma_A\frac{\mu_A(1-\delta) - \frac{1}{2}(1-\phi)\sigma_M}{\mu_A(1-\delta)[\sigma_A + \phi\sigma_M]} &\leq 1 \\
\varsigma_A &\leq \frac{\mu_A(1-\delta)[\sigma_A + \phi\sigma_M]}{\mu_A(1-\delta) - \frac{1}{2}(1-\phi)\sigma_M}
\end{aligned}$$

Proof of Proposition 4: Derivation of The Partial Separating Bound

Partial Separating Equilibrium. If A traders mix between A and the ETF, B traders stay out of the ETF so long as:

$$\text{If B} = 0\text{: } \phi \geq \frac{\beta\left(\frac{(1-\beta)\mu_B}{(1-\beta)\mu_B + \frac{1}{2}\sigma_B}\right)\left((1-\delta)\mu_A + \frac{1}{2}\sigma_A + \frac{1}{2}\sigma_M\right) - \frac{1}{2}\delta\sigma_A}{\delta\left((1-\delta)\mu_A + \frac{1}{2}\sigma_A\right) + \beta\left((1-\delta)\mu_A + \frac{1}{2}\sigma_A + \frac{1}{2}\sigma_M\right)} \tag{A1}$$

$$\text{If B} = 1\text{: } \phi \geq \frac{(1-\beta)\left(\frac{\beta\mu_B}{\beta\mu_B + \frac{1}{2}\sigma_B}\right)\left(\delta\mu_A + \frac{1}{2}\sigma_A + \frac{1}{2}\sigma_M\right) - (1-\delta)\frac{1}{2}\sigma_A}{(1-\delta)\left(\delta\mu_A + \frac{1}{2}\sigma_A\right) + (1-\beta)\left(\delta\mu_A + \frac{1}{2}\sigma_A + \frac{1}{2}\sigma_M\right)} \tag{A2}$$

Let $\sigma_M = (1 - \mu_A - \mu_B - \sigma_A - \sigma_B)$. Suppose also that A traders are mixing between the ETF and the stock. For A traders, the mixing probabilities for either the bid or the ask are:

$$\text{ETF Buy Probability (A=1): } \psi_1 = \frac{\phi\delta\mu_A\sigma_M - (1-\phi)\frac{1}{2}\sigma_A\sigma_M}{\delta\mu_A(\sigma_A+\phi\sigma_M)}$$
$$\text{ETF Sell Probability (A=0): } \psi_2 = \frac{\phi(1-\delta)\mu_A\sigma_M - (1-\phi)\frac{1}{2}\sigma_A\sigma_M}{(1-\delta)\mu_A(\sigma_A+\phi\sigma_M)}$$

The ETF bid and ask prices are:

$$(AB)_{ask} = \phi\delta\frac{\mu_A\psi_1 + \sigma_M\frac{1}{2}}{\delta\mu_A\psi_1 + \sigma_M\frac{1}{2}} + (1-\phi)\beta$$
$$(AB)_{bid} = \phi\delta\frac{\sigma_M\frac{1}{2}}{(1-\delta)\mu_A\psi_2 + \sigma_M\frac{1}{2}} + (1-\phi)\beta$$

Traders in B face the following trade-off between trading the basket at a small spread and trading the individual stock at a wide spread. The B-informed traders know the true value of B, but they share the market maker's prior about A that $P(A=1)=\delta$. Therefore they estimate the value of the ETF at $(1-\phi)B+\phi\delta$. The trade-offs that B-informed face is:

$$\text{Buy (B=1): } \phi\delta + (1-\phi) - (AB)_{ask} \leq 1 - B_{ask}$$
$$\text{Sell (B=0): } (AB)_{bid} - \phi\delta \leq B_{bid}$$

Solving for ϕ gives the result.

Proof of Proposition 5: Existence of the Pooling Equilibrium

If the conditions of Proposition 4 are violated for both securities, then there is a fully pooling equilibrium. Both traders trade the ETF, and following an ETF trade, the market maker has the following Bayesian posteriors:

$$\delta_{buy} = \delta\frac{\mu_A\psi_{A,1} + \beta\mu_B\psi_{B,1} + \frac{1}{2}\sigma_M}{\delta\mu_A\psi_{A,1} + \beta\mu_B\psi_{B,1} + \frac{1}{2}\sigma_M} \qquad \delta_{sell} = \delta\frac{\frac{1}{2}\sigma_M}{(1-\delta)\mu_A\psi_{A,2} + (1-\beta)\mu_B\psi_{B,2} + \frac{1}{2}\sigma_M}$$
$$\beta_{buy} = \beta\frac{\delta\mu_A\psi_{A,1} + \mu_B\psi_{B,1} + \frac{1}{2}\sigma_M}{\delta\mu_A\psi_{A,1} + \beta\mu_B\psi_{B,1} + \frac{1}{2}\sigma_M} \qquad \beta_{sell} = \beta\frac{\frac{1}{2}\sigma_M}{(1-\delta)\mu_A\psi_{A,2} + (1-\beta)\mu_B\psi_{B,2} + \frac{1}{2}\sigma_M}$$

Consider the case where $A=1=B$. Let $\psi_{A,1}$ be the probability that an A-informed investor buys the ETF when $A = 1$. Let $\psi_{B,1}$ be the probability that a B-informed investor buys the ETF when $B = 1$. Then it must be that $\psi_{A,1}$ and $\psi_{B,1}$ solve:

$$\phi(1-\delta_{buy}) + (1-\phi)(\beta - \beta_{buy}) = 1 - \delta \frac{\mu_A(1-\psi_A) + \frac{1}{2}\sigma_A}{\delta\mu_A(1-\psi_A) + \frac{1}{2}\sigma_A} \tag{A3}$$

$$\phi(\delta-\delta_{buy}) + (1-\phi)(1 - \beta_{buy}) = 1 - \beta \frac{\mu_B(1-\psi_B) + \frac{1}{2}\sigma_B}{\beta\mu_B(1-\psi_B) + \frac{1}{2}\sigma_B} \tag{A4}$$

where

$$\delta_{buy} = \delta \frac{\mu_A\psi_{A,1} + \beta\mu_B\psi_{B,1} + \frac{1}{2}\sigma_M}{\delta\mu_A\psi_{A,1} + \beta\mu_B\psi_{B,1} + \frac{1}{2}\sigma_M}$$

$$\beta_{buy} = \beta \frac{\delta\mu_A\psi_{A,1} + \mu_B\psi_{B,1} + \frac{1}{2}\sigma_M}{\delta\mu_A\psi_{A,1} + \beta\mu_B\psi_{B,1} + \frac{1}{2}\sigma_M}$$

The right side of Equation A3 represents the profits to A-informed investors from from trading the ETF. These profits are decreasing in ψ_A. If $\psi_B = 0$, we would have $\psi_A = \psi_1$. Since ETF profits are decreasing in ψ_B, then it must be that $\psi_A < \psi_1 < 1$.

We also know that if $\psi_A = \psi_1$, then the violation of Equation 6 in Propsition 4 would imply that B-informed investors have a profitable trading opportunity, and thus $\psi_B > 0$.

A similar logic applied to Equation A4 gives that $\psi_B < 1$ and $\psi_A > 0$.

Now since the right side of Equation A3 is decreasing in ψ_A and the left side is increasing, we have a unique ψ_A solution. Similarly, Equation A4 gives a unique ψ_B solution.

A similar argument holds for $B = 0 = A$.

Appendix B. Capital Limit

Informed investors in the model are risk neutral but face a capital limit of a single share of any security. This capital limit could be thought of as a position limit on how much investors are allowed to trade, but it can also be justified as risk aversion to an underlying common factor risk. Under this latter framework, the payoffs of assets can be redesigned as idiosyncratic payoff plus common factor risk. While Stock A previously paid a liquidating dividend from $\{0, 1\}$, the dividend

d_A with an underlying factor risk can be written as:

$$d_A = \beta_A r_M + x_A$$

where $x_A \in \{0, 1\}$. Similarly for Stock B, the liquidating dividend d_B can be written as

$$d_B = \beta_B r_M + x_B$$

where $x_B \in \{0, 1\}$. Let r_M have $E(r_M) = \mu_M$ and $Var(r_M) = \sigma^2$. Each stock-specific component is idiosyncratic, thus we have that $Cov(r_M, x_i) = 0$ for $i = A, B$.

With the standard assumptions of Ross (2013), the risk premium of idiosyncratic components of risk is zero, while common factors, such as r_M, carry a risk premium.

Under this setup, an investor who trades one share of stock A takes on β_A unites of market risk, while an investor who trades one share of stock B takes on β_B units of market risk. A single share of the ETF carries one unit of market risk. This single share of the ETF contains ϕ shares of A and $(1 - \phi)$ shares of B. If $\beta_A = 1 = \beta_B$, then an investor takes on the same amount of market risk regardless of whether they trade one share of the stock, one share of the ETF, any mixed strategy which randomizes between which asset is traded. Alternatively, if $\beta_A \neq 1$, ϕ can be redefined to ϕ^* where $\phi^* = \beta_A \phi$. In this way, an investor who trades one share of Stock A obtains β_A units of risk, while trading the ETF to take on β_A units of market risk gives $\phi\beta_A$ shares of stock A. Thus the ETF weights can be redefined so that one share of the ETF contains $\phi\beta_A$ units of stock A, and $(1 - \phi)\beta_B$ units of stock B.

Thus the normalization across securities that investors either purchase one share of the stock or one share of the ETF potentially relies on redefining the ETF weight. The ETF weight parameter, ϕ can be thought of as the literal weight of stock A in the event that $\beta_A = 1$. In the case where $\beta_A \neq 1$, ϕ is a more general trade-off between the stock and the ETF. It is a linear transformation of the raw ETF weight of stock A, scaled by the value of β. For higher β stocks, the ETF is a closer substitute, as investors are already taking on considerable factor risk when they trade the individual stock. For low β stocks, however, the ETF is a poor substitute, as they must take on considerable factor risk which they wouldn't take when trading just the stock.

As a numerical example, consider an ETF with $\beta_A = \frac{4}{3}, \beta_B = \frac{3}{4}$. Then the market portfolio has $\phi_A = \frac{3}{7}, \phi_B = \frac{4}{7}$. Suppose an investor is willing to take on \$100 worth of market risk. This investor could invest \$75 in Stock A, or \$100 in the ETF and indirectly obtain \$42.86 worth of Stock A. This ratio satisfies: $\phi_A^* = \frac{42.86}{75} = \phi_A\beta_A$. In this numerical example, $\beta_A > \beta_B$, and therefore $\phi_A^* > \phi_A$ while $\phi_B^* < \phi_B$.

Appendix C. Factor Informed Trading

The case of factor-informed traders, who know the value of some underlying factor common to all securities, is strikingly similar to the equilibria with informed traders in two independent stock-specific securities. The stock-specific dividends d_A and d_B for each security can be decomposed into:

$$d_A = \beta_A r_M + x_A$$
$$d_B = \beta_B r_M + x_B$$

where $x_A \in \{0, 1\}$, and r_M is assumed to be independent of x_A and x_B. The factor return is also assumed to have a discrete payout $r_M \in \{0, 1\}$, with the market maker's prior $M = P(r_M = 1)$. The same normalization from Appendix B is used so that the quantity of a share of stock A or a share of stock B each convey the same degree of factor risk.

In the multi-asset case of Section V, there are already informed traders with two pieces of orthogonal information. Informed traders in Stock B can be replaced with factor traders by setting σ_B, the level of noise trading in Stock B, equal to one. This forces ψ_B, the portion of orders these traders send to the ETF, to be one, just as a factor-based trader would trade.

COROLLARY 3: *In the factor-based model, the partial separating equilibrium in which only A-informed traders trade the stock and ETF is only possible if factor informed traders prefer not to* *trade the ETF at all:*

$$\textit{Bid (B=0 and B traders consider selling the ETF):} \qquad \phi\left(\frac{(1-\delta)\delta\mu_A\psi_2}{(1-\delta)\mu_A\psi_2+\frac{1}{2}\sigma_M}\right) \geq M$$

$$\textit{Ask (B=1 and B traders consider buying the ETF):} \qquad \phi\left(\frac{(1-\delta)\delta\mu_A\psi_1}{\delta\mu_A\psi_1+\frac{1}{2}\sigma_M}\right) \geq (1-M)$$

Note that compared to Corollary 2, the ETF trader's information always has full weight in the ETF: $(1-\phi)$ has been replaced with 1. In the extreme case of Corollary 3, the informational asymmetry over stock A is large enough that factor-informed traders cannot trade at the prevailing equilibrium spreads. This is natural if there is little informational asymmetry about the value of the factor return, and a large informational asymmetry about the value of Stock A.

Equilibria where the factor-informed traders also trade single stocks are not possible in a frictionless model, except in the special case of zero ETF volume. If it were the case that single stocks offered narrower spreads than the ETF, the noise traders who want factor exposure would cease trading the ETF, and only trade single stocks. With no noise trading, informed traders would be unable to profitably trade the ETF, and the ETF volume would collapse. For the vast majority of equity ETFs, the ETF spread is narrower than the weighted average spread of the underlying stocks. ETFs which violate this pattern are said to "trade rich", but the trading frictions which lead to this unusual spread behavior are outside the scope of this model.

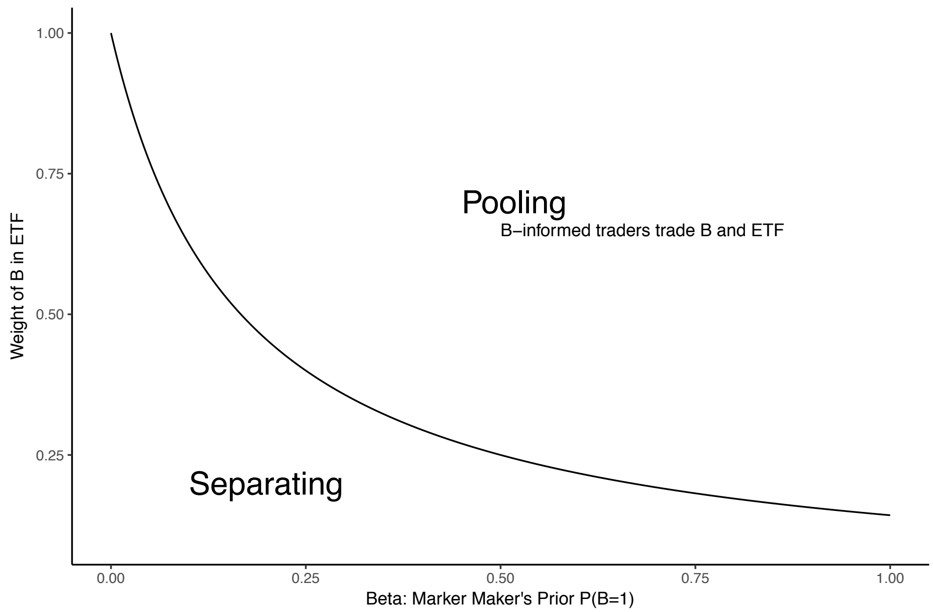
Appendix D. Pooling Equilibrium Properties

Figure 7 illustrates how trader behavior and the market maker's updates vary with changes in the ETF weights of stocks. As shown in Panel A of Figure 7, according as the stock weight in the ETF increases, investors send a higher portion of their orders to the ETF. Panel B of Figure 7 plots the difference between the market maker's prior and posterior. For stock A, this difference, $|\delta_{buy}-\delta|$, is relatively constant across a range of values of ϕ. By contrast, the difference for stock B, $|\beta_{buy}-\beta|$, changes dramatically with changes in ϕ. The difference between A and B arises from the difference in priors: the market maker is slightly more certain about the value of stock A than the value of stock B.

Changes in the prior beliefs have both a direct and indirect effect on the market maker's updating. For the direct effect, more certainty in the prior reduces the importance of new evidence. For the indirect effect, uncertainty in the stock changes spreads and therefore trading behavior. Informed traders in stocks for which the market maker has an uncertain prior face wider spreads; these wide stock spreads lead traders to trade the ETF with higher probability. Panel A of Figure 8 illustrates these effects. As the market maker's prior β about stock B increases from low levels, the B-informed traders send a higher portion of their trades to the ETF. At very high levels of β, the B-informed traders face small trading profits relative to the adverse selection from A-informed traders, and the portion of orders they send to the ETF falls off sharply. The market maker's update subsequent to an ETF trade, illustrated in Panel B of Figure 8, reflects a balance of the uncertainty in the Bayesian prior, which peaks at $\beta = \frac{1}{2}$, and the trading intensity, which peaks at $\beta \approx 0.7$.

Figure 6. Partial Separating and Pooling Equilibria. This graph depicts which equilibrium prevails for the ask quote as the ETF weight and market maker's prior of stock B change. When there are only informed traders in stock B, there is a large parameter region over which B-informed traders trade both the stock and the ETF. With price discovery in two assets, adverse selection from A-informed traders reduces the region over which the B-informed traders use a mixed strategy of randomizing between the single stock and ETF. Parameters: $\mu_A = .15 = \mu_B, \sigma_A = .05 = \sigma_B, \sigma_M = .6, \delta = .5$.

Panel A: Division of equilibria with Price Discovery in One Asset



Panel B: Division of equilibria with Price Discovery in Two Assets

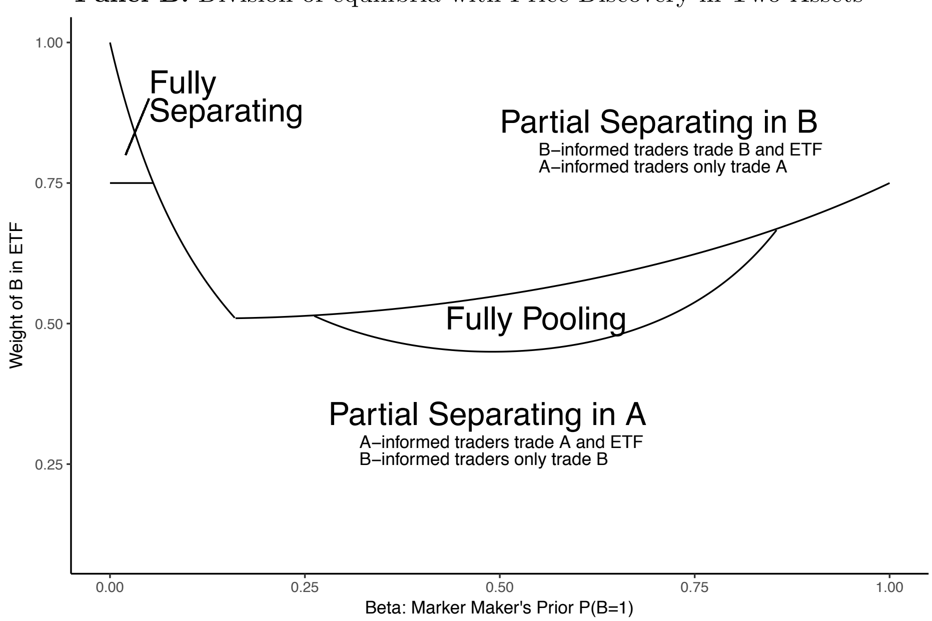
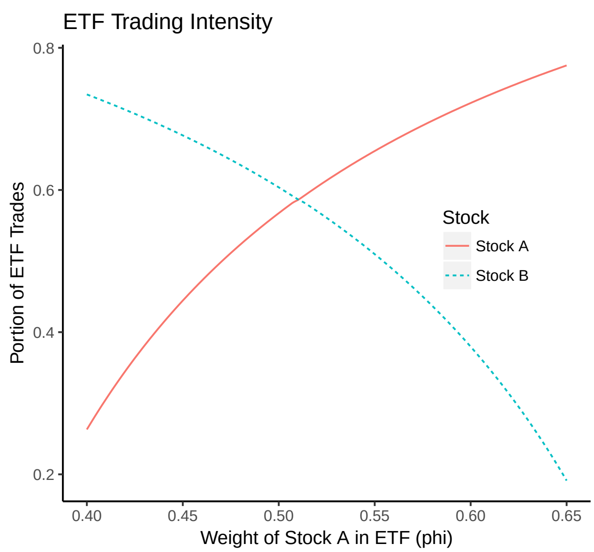
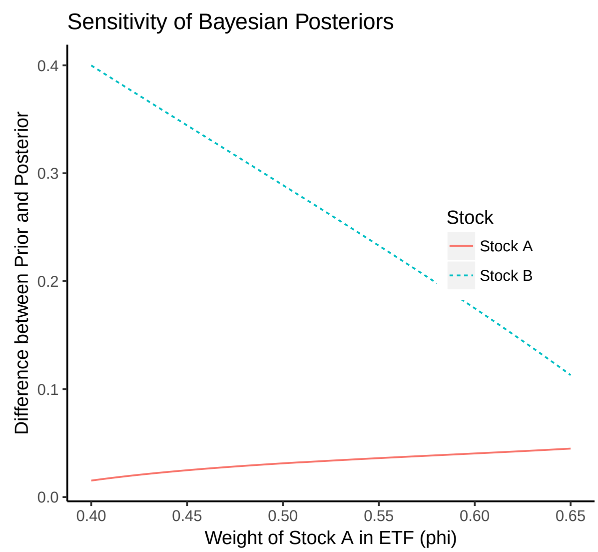


Figure 7. Pooling Equilibrium With Changes in ETF Weight. I plot how changes in ETF weight of stock A affect the trading behavior and market maker's updating for the pooling equilibrium. I plot trader behavior when $A = 1 = B$, and the updating following a buy order. The securities are symmetric in trader masses: $\mu_A = \mu_B = .15$ and $\sigma_A = \sigma_B = .15$. The priors differ, with the market maker slightly more certain about the value of stock A: $P(A = 1) = .8$ while $P(B = 1) = .75$.

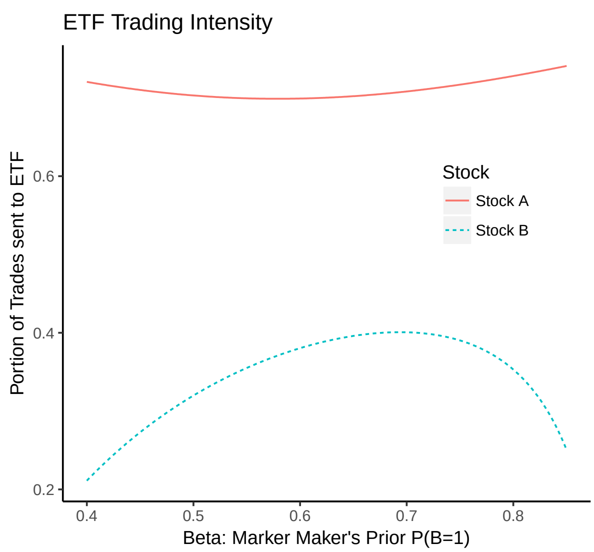


Panel A: As the ETF weight of stock A increases (x-axis), the portion of orders sent to the ETF by A-informed traders increases (solid red line) while the portion of orders sent to the ETF by B-informed traders decreases (dashed blue line). Different pieces of stock-specific information act as substitutes.

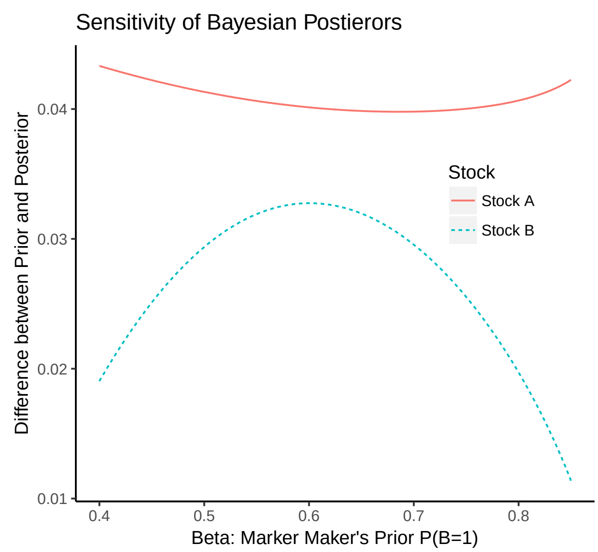


Panel B: The y-axis plots the difference between the market maker's prior and posterior $(\delta - \delta_{buy})$ for stock A, and $(\beta - \beta_{buy})$ for stock B. Increases in the ETF weight of stock A lead to a small increase in updating for stock A, and a large decrease in updating for stock B.

Figure 8. Differences in ETF Trading and Market Maker's Posterior as a Function of Prior. I plot trader behavior when $A = 1 = B$, and the updating following a buy order. The securities are symmetric in trader masses: $\mu_A = \mu_B = .15$ and $\sigma_A = \sigma_B = .15$. The ETF is 60% A (i.e. $\phi = 0.6$), and the market maker has a prior $\delta = \mathbb{P}(A = 1) = .6$.



Panel A: Increases in the market maker's prior β about the value of stock B have a non-linear effect on trading behavior. At low levels, higher levels of β increase stock-specific spreads, leading B-informed traders to send more orders to the ETF. At high levels of β, the profit B-informed traders can make in the ETF is small relative to the adverse selection they face from A-informed traders, and the portion of trades they send to the ETF decreases in β.



Panel B: The difference between the market maker's prior and posterior belief depends on a combination of the market maker's prior, the ETF weight, and trader behavior. Stock A has a higher weight, so the market maker updates more based on stock A. Quotes in Stock B change most at $\beta \approx .6$, reflecting a mix between the market maker's uncertainty (which peaks at $\beta = .5$) and the ETF-trading by B-informed traders (which peaks at $\beta \approx .7$).

Appendix E. Discrete Empirical Tests

To further explore how the stock–ETF relationship changes with the stock weight, I split my sample of stocks by their ETF weights. Assigning an indicator variable for each size group eliminates any concerns about day-to-day fluctuations in the ETF weight of each security. I define small stocks as those with a weight less than 2%, medium stocks with weight between 2% and 5%, and large stocks with a weight greater than 5%. I then estimate the following regression:

REGRESSION 4: *For stock i, ETF j, and day t:*

$$Simultaneous\ Trades_{ijt} = \alpha_0 + Size * \alpha_1 Earnings\ Date_{it} + \alpha_2 Controls_{ijt} + \epsilon_{ijt} \tag{E1}$$

$$Simultaneous\ Trades_{ijt} = \alpha_0 + Size * \alpha_1 Abs\ Return_{it} + \alpha_2 Controls_{itj} + \epsilon_{ijt} \tag{E2}$$

Results are presented in Table XI. Small stocks see fewer trades on earnings dates, while large stocks see far more. This is consistent with the idea that investors with information about small stocks face adverse selection in the ETF. During earnings dates, the adverse selection in the ETF would outweigh any benefit from trading the ETF based on their small-stock information. Across returns, however, there is no evidence of exclusion, with larger absolute returns leading to more simultaneous trades for each of the three stock categories.

To investigate how simultaneous trading activity varies with the stock-specific return, I further sub-divide my sample based with an indicator on the size of each return:

REGRESSION 5: *For stock i, ETF j, and day t:*

$$Simultaneous\ Trades_{ijt} = \alpha_0 + Size * \alpha_1 Largest\ X\ Abs\ Return_{it} + \alpha_2 Controls_{itj} + \epsilon_{ijt} \tag{E3}$$

Largest X Abs Return is an indicator that takes the value of 1 on the dates for which each stock has its X% most positive and X% most negative intraday returns. Controls include a fixed effect for each ETF as well as a similarly defined Largest X indicator which takes the value of 1 on the dates for which each ETF has its X% most positive and X% most negative intraday returns.

Results are presented in Table XII. Across all three size categorizations of stocks, stocks have more stock–ETF simultaneous trades on dates with large absolute returns. This pattern increases in the level of the return: as the indicator on returns selects more extreme returns, the estimated coefficient increases. Large stocks have the highest level of simultaneous trades, but all three size categories of stocks present the same pattern that more extreme returns lead to more trade-both behavior by investors.

When short selling assets, investors must account for the short-sale costs involved. Huang et al. (2021) find an asymmetry between long and short positions in an ETF, and argue that the lower shorting costs of ETFs drive this result. To test for similar asymmetric effects in my setting, I re-estimate Regression 1 and Regression 5 restricted to positive or negative stock-specific returns. Results are presented in Table XIII. When investors compare trading both the stock and ETF against trading only the stock, the reduced shorting costs of the ETF increase the attractiveness of the stock-ETF option. Consistent with this intuition, estimates in Table XIII are stronger for days with a negative stock-specific return.

Table XI Estimation of Regression 4: Size Comparison

This table reports estimates of Regression 4, which estimates the effect of changes in stock-specific information across different ETF weights. Earnings Date is an indicator which takes the value 1 for stocks which announce earnings before the day's trading session. Abs Return is the absolute value of the intraday return, measured as a percentage. I categorize small stocks as those with a weight less than 2%, medium stocks with weight between 2% and 5%, and large stocks with a weight greater than 5%. The sample is all stocks listed in the S&P 500 index at any point August 1, 2015 to December 31, 2018, and the 552 ETFs selected for having at least 50% of their holdings invested in said stocks and trading on at least ten days. The frequency of observations is daily. I include a fixed effect for each ETF and cluster standard errors by ETF.

	Simultaneous Trades	
	(1)	(2)
Small Stocks	−580.893***	−386.708***
	(154.377)	(110.974)
Medium Stocks	−342.741***	−191.765***
	(123.170)	(65.721)
Earnings Date∗Small Stock	64.997*	
	(34.144)	
Earnings Date∗Medium Stock	112.505***	
	(28.602)	
Earnings Date∗Large Stock	483.632***	
	(181.780)	
Abs Return∗Small Stock		46.216**
		(22.115)
Abs Return∗Medium Stock		97.383***
		(24.368)
Abs Return∗Large Stock		297.822***
		(88.609)
Observations	34,068,858	34,068,205
R^2	0.286	0.288
Residual Std. Error	1,482	1,480
Standard Errors Clustered by ETF:	*p<0.1; **p<0.05; ***p<0.01	

Table XII : Estimation of Regression 5 - Return Comparison

This table reports estimates of Regression 5. For each stock, largest $X\%$ Abs Return is an indicator which takes the value one on days for which the intraday return is among the most positive $X\%$ or most negative $X\%$ of returns for that stock. I include an equivalently defined daily indicator on the ETF return for whether the ETF return is among the most positive $X\%$ or most negative $X\%$. Small stocks are stocks with less than 2% ETF weight, medium stocks are 2% to 5%, and large stocks have greater than 5% weight. The sample is all stocks listed in the S&P 500 index at any point August 1, 2015 to December 31, 2018, and the 552 ETFs selected for having at least 50% of their holdings invested in said stocks and trading on at least ten days. The frequency of observations is daily. I include a fixed effect for each ETF and cluster standard errors by ETF.

Panel A: Regression Estimates with Small Stocks

	Simultaneous Trades			
	(1)	(2)	(3)	(4)
Largest 1% Abs Return	216.338**			
	(108.935)			
Largest 5% Abs Return		151.619**		
		(70.637)		
Largest 10% Abs Return			126.839**	
			(56.775)	
Largest 20% Abs Return				96.287**
				(41.853)
Largest X% ETF return	x	x	x	x
Observations	32,078,562	32,078,562	32,078,562	32,078,562
R^2	0.314	0.314	0.313	0.312
Residual Std. Error	1,348	1,348	1,348	1,349

Panel B: Regression Estimates with Medium Stocks

	Simultaneous Trades			
	(1)	(2)	(3)	(4)
Largest 1% Abs Return	477.496***			
	(109.598)			
Largest 5% Abs Return		342.931***		
		(82.992)		
Largest 10% Abs Return			269.595***	
			(61.976)	
Largest 20% Abs Return				193.370***
				(42.330)
Largest X% ETF return	x	x	x	x
Observations	1,667,497	1,667,497	1,667,497	1,667,497
R^2	0.350	0.350	0.349	0.349
Residual Std. Error	2,382	2,383	2,383	2,385

Table XII: (continued from previous page)

Panel C: Regression Estimates with Large Stocks

	Simultaneous Trades			
	(1)	(2)	(3)	(4)
Largest 1% Abs Return	1,186.083***			
	(358.117)			
Largest 5% Abs Return		778.781***		
		(213.962)		
Largest 10% Abs Return			126.839**	
			(56.775)	
Largest 20% Abs Return				448.901***
				(125.010)
Largest X% ETF return	x	x	x	x
Observations	312,642	312,642	312,642	312,642
R^2	0.335	0.336	0.335	0.333
Residual Std. Error	3,310	3,310	3,316	3,316

Table XIII Estimation of Regression 1: Directional Comparison

This table reports estimates of Regression 1 and Regression 5 restricted to positive or negative subsamples. Columns (1) and (3) are restricted to dates where the stock-specific return is positive, while Columns (2) and (4) are restricted to dates where the stock-specific return is negative. Effects are stronger when the return is negative, which reflects the intuition from Huang et al. (2021) that ETF shorting costs are cheaper than stock shorting costs. Abs Return is the absolute value of the intraday return, measured as a percentage. For each stock, largest 5% Abs Return is an indicator which takes the value one on days for which the intraday return is among the most positive 5% or most negative 5% of returns for that stock.The sample is all stocks listed in the S&P 500 index at any point August 1, 2015 to December 31, 2018, and the 552 ETFs selected for having at least 50% of their holdings invested in said stocks and trading on at least ten days. The frequency of observations is daily. I include a fixed effect for each ETF and cluster standard errors by ETF.

	Dependent variable: Simultaneous Trades			
Return:	Positive	Negative	Positive	Negative
	(1)	(2)	(3)	(4)
Weight	89.490***	102.995***	83.194**	98.507**
	(27.849)	(35.073)	(32.442)	(42.023)
Largest 5% Return	85.496*	143.338*		
	(47.841)	(83.227)		
Weight∗Largest 5% Return	67.250***	97.194***		
	(18.167)	(27.192)		
Absolute Return			18.670*	22.934*
			(9.994)	(12.405)
Weight∗Absolute Return			18.469***	24.538***
			(4.903)	(6.848)
Largest 5% ETF Return	X	X		
Weight∗Largest 5% ETF Ret.	X	X		
Absolute ETF Return			X	X
Observations	17,689,036	16,137,504	17,684,068	16,132,754
R^2	0.321	0.274	0.323	0.277
Residual Std. Error	1,248	1,693	1,246	1,690

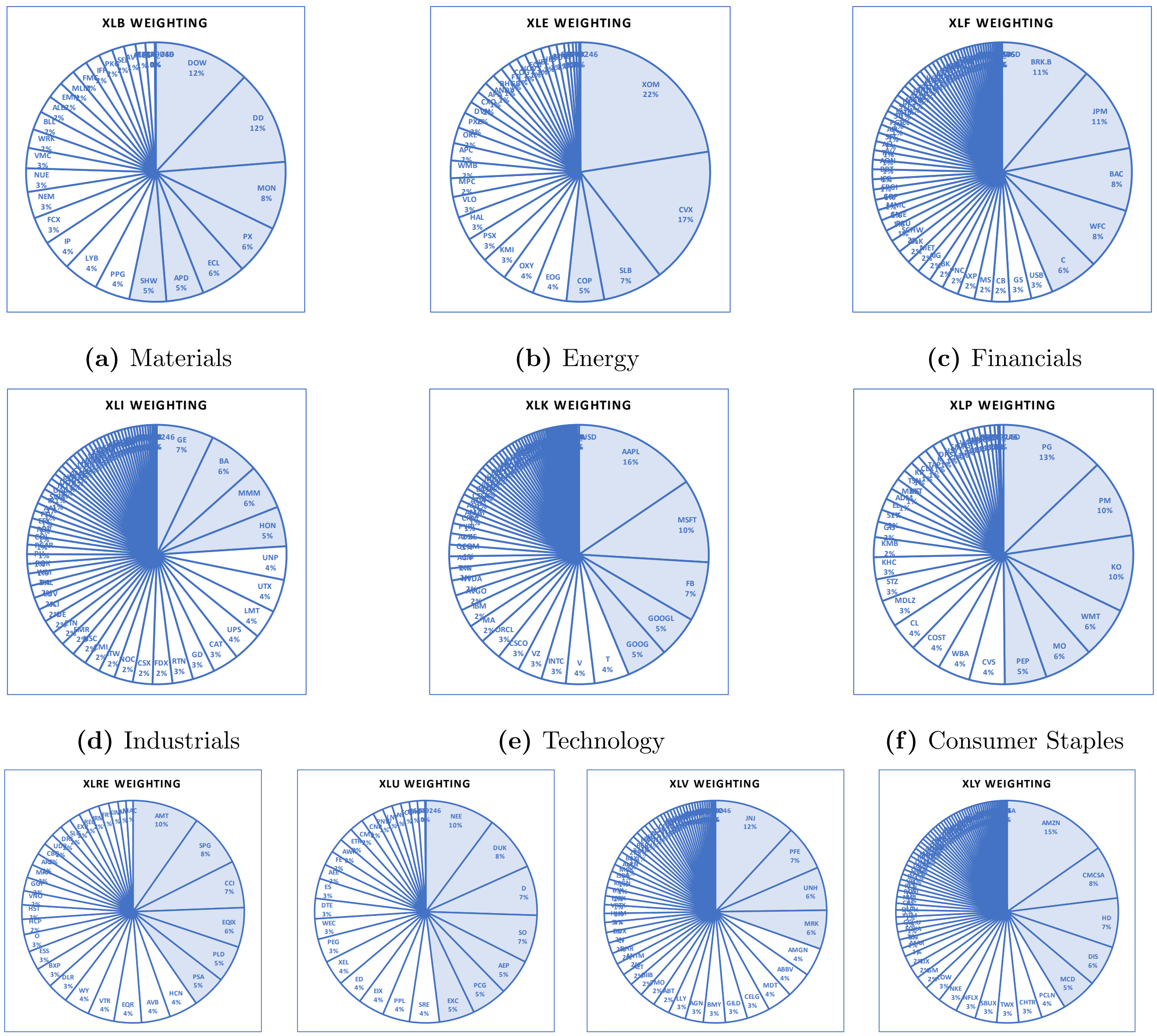
Standard Errors Clustered by ETF: *p<0.1; **p<0.05; ***p<0.01

Appendix F. SPDR and Sector SPDR Robustness Tests

My sample of ETFs is SPDR and the ten Sector SPDR ETFs from State Street. The Sector SPDRs divide the stocks of the S&P500 into ten GICS industry groups,[18] and have the advantage of being very liquid, fairly concentrated, and representative of a broad set of securities. Within each ETF, constituents are weighted according to their market cap. As a result, the ETFs have fairly concentrated holdings, as depicted in Figure 9, with the stocks with more than 5% weight comprising between 25% and 50% of each ETF.

The SPDR ETFs are extremely liquid. SPDR itself has the largest daily trading volume of any exchange-traded product. All of the Sector SPDRs are in the top 100 most heavily traded ETFs, with 6 in the top 25 most traded ETFs. With $30 billion per day in trading volume, the ETFs in my sample represent over 30% of total ETF trading volume. All the stocks they own are also domestically listed and traded on the main US equities exchanges. Trading volume of these ETFs is high even compared to the very liquid underlying stocks. For example, the Energy SPDR (XLE) has an average daily trading volume of around 20 million shares, at a price of $65 per share. As listed in Panel 9b, 17% of the holdings of XLE are Chevron stock. This means that when investors buy or sell these 20 million ETF shares, they are indirectly trading claims to $218 million worth of Chevron stock. The average daily volume for Chevron stock is $800 million, so the amount of Chevron that changes hands within the XLE basket is equal in size to 30% of the daily volume of Chevron stock.

[18] The GICS Groups are: Financials, Energy, Health Care, Consumer Discretionary, Consumer Staples, Industrials, Materials, Real Estate, Technology, and Utilities. During my sample period, two reclassifications of groups occurred. In September 2016, the Real Estate Sector SPDR (XLRE) was created from stocks previously categorized as Financials. In October 2018, the Communications Sector SPDR (XLC) was created from stocks previously categorized as Technologies or Consumer Discretionary. Since XLC has only three months of transaction data in my sample, I exclude it from the analysis.



(a) Materials **(b)** Energy **(c)** Financials

(d) Industrials **(e)** Technology **(f)** Consumer Staples

(g) Real Estate **(h)** Utilities **(i)** Health Care **(j)** Consumer Discr.

Figure 9. Holdings of the Sector SPDRs. Stocks which comprise more than 5% of the ETF holdings are highlighted in light blue. The Sector SPDRs are highly concentrated. Exxon Mobile, for example, comprises 22% of the holdings of the Energy ETF, and just four stocks comprise over half the holdings of XLE.

Table XIV: Summary Statistics on Securities

(a) Panel A: Stock Summary Statistics
My sample is comprised of the stocks of the S&P 500 Index. During my sample period from August 1, 2015 to December 31, 2018, there are 860 trading days.

Statistic	Mean	St. Dev.	Min	Max
Daily Simultaneous Trades	102	321	0	13,665
ETF Weight	1%	1.6%	0.1%	25%
Daily ETF Orders	81,052	149,876	480	2,035,648
Daily Stock Orders	20,216	21,161	12	748,384
Daily Return	0.1%	1.7%	−41%	71%

(b) Panel B: Sector SPDR Summary Statistics The ten Sector SPDRs divide the S&P 500 index into portfolios based on their GICS Industry Code. I categorize small stocks as stocks with an ETF-constituent weight of $< 2\%$, medium stocks as stocks with an ETF-constituent weight between 2% and 5%, and large stocks as stocks with an ETF-constituent weight greater than 5%.

ETF	Industry	Small Stocks	Medium Stocks	Large Stocks	Mean ETF Return (%)	Std Dev. (Return %)
XLV	Health Care	46	11	4	0.028	0.71
XLI	Industrial	53	12	4	0.027	0.76
XLY	Consumer Disc.	55	6	3	0.041	0.70
XLK	Technology	54	10	4	0.030	0.83
XLP	Consumer Staples	19	9	4	0.010	0.64
XLU	Utilities	7	16	6	0.042	0.78
XLF	Financials	40	6	3	-0.008	0.85
XLRE	Real Estate	14	13	4	0.031	0.79
XLB	Materials	10	11	4	0.007	0.77
XLE	Energies	15	12	3	-0.008	0.93

Appendix G. Placebo Tests Tests

In this section, I analyze cross market activity between the same stocks of the S&P 500, and a set of equal-weighted ETFs from Investco. For each of the State Street SPDR ETFs, there is an equivalent Investco Equal Weight ETF which holds the same exact stocks. The Investco ETFs, however, give each stock an equal weight in the ETF rather than value weights.

I measure cross market activity between stocks and the equal-weight ETF. When I look at how cross market activity changes across stocks, however, I use the market value weights of each stock. Thus I expect a null result: large-cap stocks should not have any more cross market activity with Investco ETFs than small-cap stocks, given that the Investco ETFs hold all stocks in equal proportion. Table XV shows the estimation of Regression 1. Coefficient estimates are almost all near zero, and there is no difference between large and small stocks.

One drawback of the Investco Equal Weight ETFs is that they have low daily trading volume. I therefore conduct a second test, this time using the Sector SPDRs with improper pairings. By improper pairings, I match a stock from sector i with the ETF for sector j, using the sector weight of stock i. As before, I expect a null result: if a stock like Exxon announces earnings or has a large return, investors with Exxon-specific information should not simultaneously trade Exxon and an unrelated Sector SPDR, like the Financials or Health Care ETF.

Table XVI presents the results of this comparison. The coefficient estimate for earnings dates is negative and insignificant. The coefficient estimates for the weight-return and weight-volatility interaction terms are positive, but around 5 times smaller than the previous estimates.

Table XV : Estimation of Regression 1 with Equal Weight ETFs (Placebo)
This table reports estimates of Regression 1 on a placebo sample. stock–ETF simultaneous trading activity is measured with a set of Investco equal-weight ETFs. Sector SPDR Weight is the (value-weighted) ETF weight of each stock in the State Street SPDR. The sample is the ten Investco Sector ETFs and their stock constituents from August 1, 2015 to December 31, 2018. The frequency of observations is daily. I include a fixed effect for each ETF and cluster standard errors by ETF.

	Simultaenous Trades			
	(1)	(2)	(3)	(4)
Sector SPDR Weight	0.501	0.869	0.558	0.850
	(0.466)	(0.640)	(0.439)	(0.615)
Earnings Date	-0.578^{**}	-0.469^{***}		
	(0.258)	(0.161)		
Abs Return			0.821^{**}	0.176^{*}
			(0.368)	(0.099)
Sector SPDR Weight∗Earnings Date	0.105^{**}	0.096^{***}		
	(0.048)	(0.036)		
Sector SPDR Weight∗ Abs Return			-0.021	0.106
			(0.056)	(0.111)
Abs ETF Return		3.621^{*}		3.506^{*}
		(1.860)		(1.802)
Sector SPDR Weight∗Abs ETF Return		-0.474^{**}		-0.567^{*}
		(0.241)		(0.331)
Observations	969,378	969,378	969,378	969,366
R^2	0.100	0.125	0.100	0.126
Residual Std. Error	13.079	12.890	13.079	12.885
Standard Errors Clustered by ETF.		$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$		

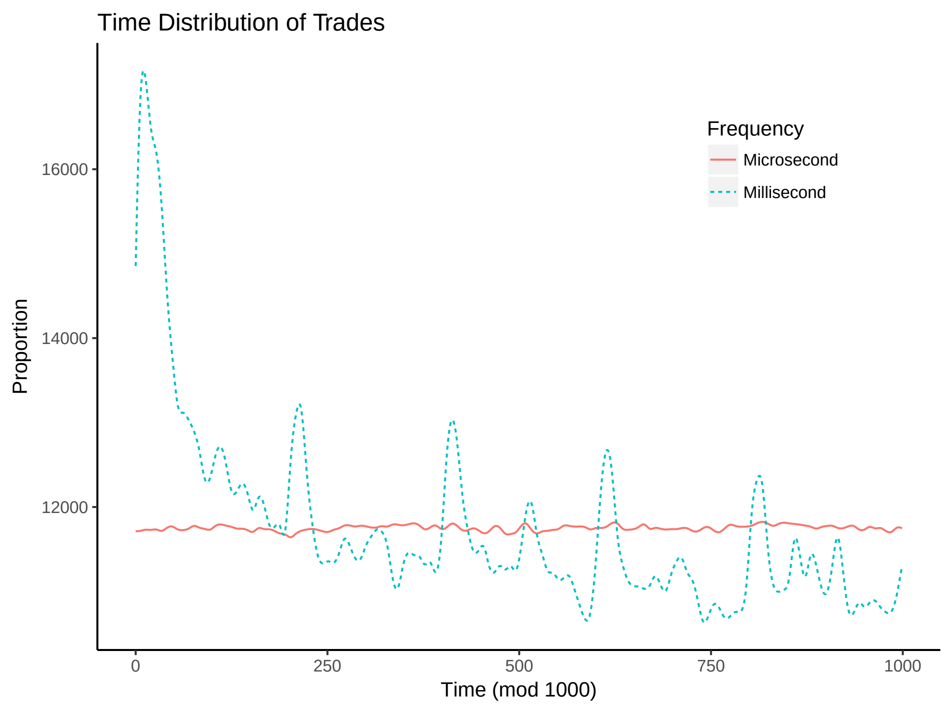
Table XVI : Estimation of Regression 1 with Improper SPDR Pairings (Placebo)
This table reports estimates of Regression 1 on a placebo sample. stock–ETF simultaneous trading activity is measured between each sector SPDR and any S&P 500 Constituents which are not in that sector. Weight is the value weight of each stock in its industry sector. The sample is the ten Sector SPDR ETFs and the S&P stock constituents from August 1, 2015 to December 31, 2018. The frequency of observations is daily. I include a fixed effect for each ETF and cluster standard errors by ETF.

	Dependent variable: Simultaneous Trades	
	(1)	(2)
Weight	3.239***	3.030***
	(0.644)	(0.631)
Earnings Date	−0.234	
	(0.425)	
Abs Return		0.152
		(0.563)
Weight∗ Earnings Date	−0.442	
	(0.308)	
Weight∗ Abs Return		0.844**
		(0.394)
Abs ETF Return	5.149***	5.188***
	(0.939)	(1.039)
Weight∗ Abs ETF Return	2.312*	1.599
	(1.326)	(1.351)
Observations	4,366,305	4,366,251
R^2	0.172	0.186
Adjusted R^2	0.172	0.186
Residual Std. Error	45.107	44.721
Standard Errors Clustered by ETF.		*p<0.1; **p<0.05; ***p<0.01

Appendix H. Trade Timing Tests

In this section, I verify that the results are not due to patterns in the timing of trading activity. Trading activity is not uniform throughout the day. Hasbrouck and Saar (2013), for example, note that there is a large spike in trading activity in the first few milliseconds of each second, suggesting algorithmic trading at a one-second frequency. Higher levels of algorithmic trading would lead to higher levels of trading in the first few microseconds of each second.

Figure 10. Time Distribution of Trades. I plot clock-time periodicities of trades for a subset of the data: XLE and the top 10 underlying stocks for September, 2016. Consider a timestamp in the form HH:MM:SS.mmm$\mu\mu\mu$. The three blue digits are the milliseconds, while the three red digits are the microseconds. There is a clear spike in trades in the first few milliseconds of each second (dashed blue line). Hasbrouck and Saar (2013) argue this could be algorithmic trading activity. The distribution of trades at the microsecond level (solid red line), however, is uniform.



I plot the distribution of trades at the millisecond and microsecond level in Figure 10. At the microsecond level, trade distribution is uniform. To confirm that my results are not driven by some form of periodic algorithmic trading, I re-estimate Regression 1 on a restricted subsample. From the sample of all trades, I eliminate trades occurring in the first 200 milliseconds of each second. This leaves trades occurring only between $(200, 999)$ of each millisecond to avoid any spikes in activity around 0 microseconds. Results for this regression are presented in Table XVII and are similar to the previous estimation of Regression 1.

Table XVII: **Robustness Check of Regression 1- Time Restricted Sub-Sample**
This table reports estimates of Regression 1 on a time restricted subsample. To control for changes in algorithmic trading, I exclude all trades which occur in the first 200 milliseconds of each second. I measure stock–ETF simultaneous trades on the trades which remain. Earnings Date is an indicator which takes the value 1 for stocks which announce earnings before the day's trading session. Abs Return is the absolute value of the intraday return, measured as a percentage. The sample is SPDR and the ten Sector SPDR ETFs and their stock constituents from August 1, 2015 to December 31, 2018. The frequency of observations is daily. I include a fixed effect for each ETF and cluster standard errors by ETF.

	Simultaneous Trades	
	(1)	(2)
Weight	16.064***	15.110***
	(5.882)	(5.468)
Earnings Date	−1.099	
	(0.909)	
Weight∗Earnings Date	2.400***	
	(0.640)	
Abs Return		1.367**
		(0.583)
Weight∗Abs Return		3.738**
		(1.754)
Abs ETF Return	37.335***	36.548***
	(14.203)	(13.908)
Weight∗Abs ETF Return	4.089	0.898
	(3.179)	(3.749)
Observations	873,244	873,226
R^2	0.166	0.169
Adjusted R^2	0.166	0.169
Residual Std. Error	116.811	116.553
Standard Errors Clustered by ETF.	*p<0.1; **p<0.05; ***p<0.01	

For each stock-ETF simultaneous trade, I check for simultaneous trades in additional assets. I consider two different possibilities for these additional assets. For a stock i in sector j, the first variant is to check for simultaneous trades with SPY, the S&P 500 ETF. The second variant is to check for simultaneous trades with each of the k other stocks in sector j.

Investors trade stocks and sector ETFs simultaneously. Investors also trade stocks and SPY simultaneously. Finally, investors trade sector ETFs and SPY simultaneously. This joint trading behavior of investors motivates the first variant of my triple-trades test: examining stock–Sector ETF–SPY simultaneous trades. This test analyzes the difference between stocks which are large in their sector, but small in SPY. Table XVIII highlights some of these stocks.

Table XVIII: **Differences Between Sector Weight and SPDR Weight**. This table highlights some of the differences between a stock's sector ETF weight and SPY Weight. Large companies in a small sector, such as American Tower (AMT) in the Real Estate Sector SPDR (XLRE) or Linde (LIN) in the Materials Sector SPDR (XLB), have a large weight within their sector, but a small weight in the SPDR ETF. Conversely, large companies in a large sector, such as Apple (AAPL) in the Technology Sector SPDR (XLK) or Amazon (AMZN) in the Consumer Discretionary Sector (XLY), have a large weight in both their sector ETF and in SPY.

Stock	Sector Wt.	Sector	SPY Weight
AAPL	20%	Technology	5%
AMZN	23%	Consumer Discr	3%
AMT	13%	Real Estate	0.4%
LIN	16%	Materials	0.4%

For stock i and ETF j on date t, I define *Simultaneous with* SPY_{ijt} as the total number of stock trades which have both a trade in the Sector SPDR ETF and SPY. I run the following regression:

REGRESSION 6: *For stock i, ETF j, and day t:*

$$\begin{aligned} Simultaneous\ with\ SPY_{ijt} = \alpha_0 &+ \alpha_1 Earnings\ Date_{it} + \alpha_2 Weight_{ij} \\ &+ \alpha_3 SPY\ Weight_{ij} + \alpha_4 Weight_{ij} * Earnings\ Date_{it} \\ &+ \alpha_5 SPY\ Weight_{ij} * Earnings\ Date_{it} + \alpha_6 Controls_{ijt} + \epsilon_{ijt} \end{aligned} \tag{H1}$$

$$\begin{aligned} Simultaneous\ with\ SPY_{ijt} = \alpha_0 &+ \alpha_1 Abs\ Return_{it} + \alpha_2 Weight_{ij} \\ &+ \alpha_3 SPY\ Weight_{ij} + \alpha_4 Weight_{ij} * Abs\ Return_{it} \\ &+ \alpha_5 SPY\ Weight_{ij} * Abs\ Return_{it} + \alpha_6 Controls_{itj} + \epsilon_{ijt} \end{aligned} \tag{H2}$$

Earnings Date is an indicator that takes the value of 1 on the trading day after a company releases earnings. Abs Return is the absolute value of the intraday return of a stock. Controls include a fixed effect for each ETF, the ETF return, and an interaction between stock weight and the ETF return.

Theory predicts a positive value for α_5. Trading both the Sector SPDR and SPY only makes sense when the stock has a substantial weight in SPY. For example, a stock like Apple comprises over 20% of the Technology Sector SPDR ETF, but also comprises a substantial 5% of SPDR. In contrast, a large stock in a small sector, like American Tower Corporation, comprises 14% of the Real Estate Sector SPDR, but less than 0.4% of SPDR. In the online Appendix, individual regressions estimated for each sector ETF confirm that investors in each of these stocks place simultaneous stock-Sector ETF trades. The stock-Sector-SPDR triple, however, is only a reasonable trade for investors informed about the largest stocks in the market. Only for these stocks do investors have a sufficient informational advantage in SPDR as well as the sector. Accordingly, in the estimation of Regression 6, coefficients α_2 and α_4 should not be large or significant: the sector SPDR weight of an ETF offers little explanatory power compared to the SPY weight.

Results of Regression 6 are presented in Table XIX. Coefficients for both the α_3 (SPY Weight) and α_5 (SPY Weight$*$Information interaction) are positive and statistically significant. In contrast, the terms α_2 (Sector Weight) and α_4 (Sector Weight$*$Information interaction) are small and not significant. This is consistent with the idea that investors in large stocks within the S&P 500 can trade both the sector ETF and SPY, while investors in small stocks in SPY which are nonetheless large stocks within their sector should not trade both the sector ETF and SPY.

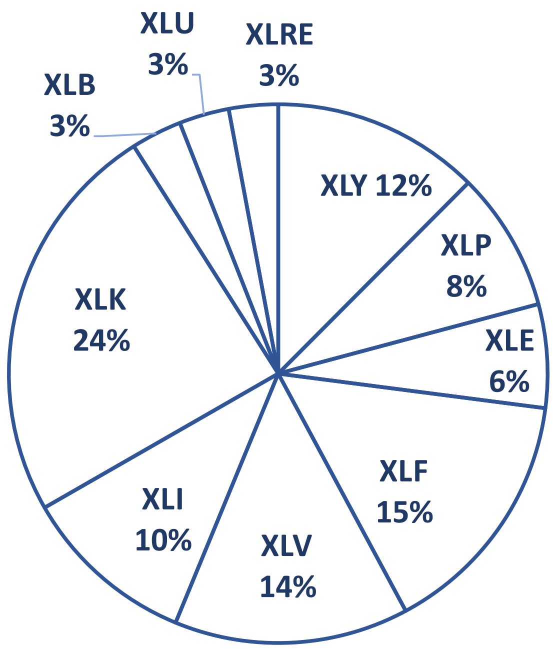
Table XIX: Estimation of Regression 6

This table reports estimates of Regression 6, which estimates the effect of changes in stock-specific information on simultaneous trades between the triple of a stock, sector SPDR, and SPY. Weight is the stock weight in its Sector SPDR ETF, while SPY weight is the weight of the stock in the S&P 500 ETF, SPY. I consider two different measures of stock-specific information: earnings dates and absolute value of the return. Earnings After is an indicator which takes the value 1 for stocks which announce earnings either before the day's trading session, or on the previous evening after the market close. Earnings Before is an indicator which takes the value for stocks which announce earnings either after the day's trading session or subsequent morning. Abs Return is the absolute value of the intraday return, measured as a percentage. The sample is SPDR and the ten Sector SPDR ETFs and their stock constituents from August 1, 2015 to December 31, 2018. The frequency of observations is daily. I include a fixed effect for each ETF and cluster standard errors by ETF.

	Simultaneous with SPY		
	(1)	(2)	(3)
Weight	−0.162	−0.554**	−0.550**
	(0.112)	(0.223)	(0.222)
Abs Return	−0.087		
	(0.260)		
Earnings After		−0.450**	
		(0.219)	
Earnings Before			0.187
			(0.284)
SPY Weight	12.685***	27.954***	27.898***
	(1.306)	(3.983)	(3.984)
Weight∗Abs Ret	−0.322*		
	(0.165)		
SPY Weight∗Abs Ret	14.737***		
	(2.957)		
Weight∗Earnings After		−0.057	
		(0.077)	
SPY Weight∗Earnings After		1.758	
		(1.910)	
Weight∗Earnings Before			−0.341
			(0.236)
SPY Weight∗Earnings Before			5.607***
			(2.040)
Observations	440,826	440,834	440,834
R^2	0.175	0.127	0.127
Adjusted R^2	0.175	0.127	0.127
Residual Std. Error (df = 440819)	20.958	21.552	21.552
Note:			*p<0.1; **p<0.05; ***p<0.01

Appendix I. SPDR and Sector SPDRs

Figure 11. SPDR Sector Weights. The stocks of SPDR are divided into sector SPDR ETFs according to their GICS Industry Classification codes. This chart shows the weight of each sector in the S&P 500.



The stocks of S&P 500 are divided into Sector SPDR ETFs according to their GICS Industry Classification codes. The SPDR ETF could therefore be thought of as a portfolio of the different Sector SPDRs. Figure 11 shows the weight of each sector in the S&P 500.

Just as investors with stock-specific information should trade ETFs, investors with Sector SPDR information should trade SPDR. I re-estimate Regression 1 using the Sector SPDR pairings with SPDR. While there are not earnings dates for sector SPDR, I can use the absolute value of the intraday sector return as a measure of information asymmetry.

Results are in Table XX. Larger absolute sector returns lead to a significant increase in simultaneous trades. This is consistent with the idea that investors with sector-specific information who are looking for additional liquidity in a sector also trade SPDR itself. In the estimation of Equation H2, larger weight sectors also see a significant increase in cross market activity. The weight-return interaction is positive, but not significant after controlling for the index return.

Table XX: Estimation of Regression 1 with Sector ETFs and SPDR

This table reports estimates of Regression 1 using Sector-SPDR pairings with the SPDR index. Abs Sector Return measures the intraday return of a Sector SPDR. I also control for the return on the SPDR index, and control for a weight-return interaction. The sample is SPDR and the ten Sector SPDR ETFs from August 1, 2015 to December 31, 2018. The frequency of observations is daily.

	Dependent variable: Simultaneous Trades	
	S	
	(1)	(2)
Weight	45.652***	55.705***
	(5.717)	(5.221)
Abs Sector Return	735.027***	411.890***
	(114.259)	(127.013)
SPY Return		918.593***
		(114.557)
Weight∗Abs Sector Return	61.662***	7.574
	(8.996)	(12.866)
Weight∗SPY Return		52.398***
		(16.509)
Constant	475.265***	181.741***
	(73.155)	(69.840)
Observations	8,257	8,257
R^2	0.154	0.199
Adjusted R^2	0.154	0.198
Residual Std. Error	2,789.534 (df = 8253)	2,715.770 (df = 8251)
Note:		*p<0.1; **p<0.05; ***p<0.01