Foxes in the Hen-House: Single Stock Traders Hiding in ETF Order Flow

By

Thomas Ernst

SUBMITTED TO THE SLOAN SCHOOL OF MANAGEMENT IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN MANAGEMENT RESEARCH

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2019

Signature of Author: Signature redacted

Department of Management
May 3, 2019

Certified by: Signature redacted

Haoxiang Zhu
Associate Professor Of Finance
Thesis Supervisor

Accepted by: Signature redacted

Ezra Zuckerman Sivan
Alvin J. Siteman (1948) Professor of Entrepreneurship and Strategy
Deputy Dean
Faculty Chair, MIT Sloan PhD Program

Foxes in the Hen-House:

Single-Stock Traders Hiding in ETF Order Flow

by
Thomas Ernst

SUBMITTED TO THE SLOAN SCHOOL OF MANAGEMENT ON MAY 3, 3019

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN MANAGEMENT RESEARCH

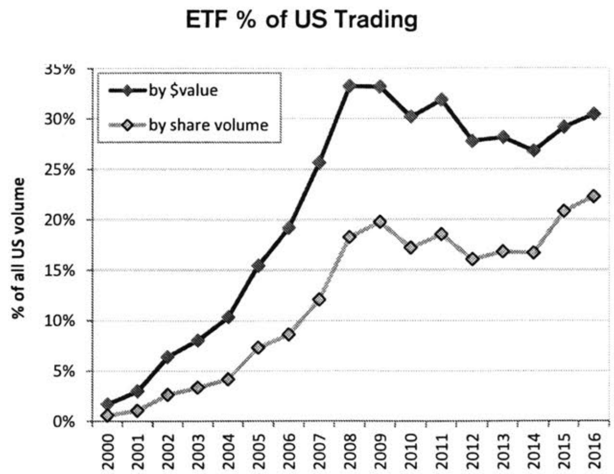
ABSTRACT

Conventional wisdom suggests that exchange traded funds (ETFs) protect uninformed investors from informed traders by diversifying holdings across a basket of securities. Contrary to this belief, I develop a theoretical model where investors with stock-specific information do trade in ETFs whenever the stock's weighting in the ETF is high or whenever the information asymmetry is large. As a result, ETF trades can allow price discovery about specific underlying stocks. High-frequency evidence from Sector SPDR ETFs supports the predictions of my model. I conclude that ETFs, especially narrowly constructed ETFs, are far from immune to single-stock informed trading.

Thesis Supervisor: Haoxiang Zhu

Title: Professor of Finance

I. Introduction



Source: Credit Suisse Trading Strategy

After a decade of rapid growth, exchange traded funds (ETFs) now comprise between 25% and 40% of total US exchange trading volume (Wigglesworth, 2017). The ascendancy of ETFs has raised concerns over their impact on the price discovery process. One concern is that single stock price discovery is harmed if uninformed investors move from trading individual stocks to trading fixed baskets of stocks. Single-stock informed traders would have no noise traders to transact against. A related concern is that ETF trading could lead to excessive co-movement. With high levels of ETF trading, liquidity shocks to the ETF could lead to arbitrage activity that leads to 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 would never trade ETFs.

This paper is the first to model theoretically, and document empirically, that investors will trade ETFs based on stock-specific information, and this interplay generates new conclusions about the link between stocks and the ETFs that include them. In my framework, investors have freedom over which assets they trade. While their information is stock-specific, trading both the stock and an ETF may reduce the market impact of their trade. Trading both assets can therefore be more profitable than just trading the stock, especially when stock-specific bid ask spreads are wide and the stock's ETF weight is high. With this trading behavior in mind, both concerns from the previous paragraph are attenuated. First, noise traders in ETFs will face adverse selection from stock-specific information. Even if stock-specific liquidity is low, traders with stock-specific information will have profitable trading opportunities so long as they can use their information to trade the ETF. Second, since stock-specific investors trade the ETF, trades in the ETF can have an asymmetric impact on the underlying stocks. Instead of uniform co-movement, an ETF trade may lead to a large quote change on some stocks and no change for others.

Formally, the model is an extension of Glosten and Milgrom (1985) with three assets: stock A, stock B, and an ETF 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, and the price of B remains fixed. The informed investors know the value of A exactly, and face a strategic choice over trading stock A or the ETF. While the ETF holds only a fraction of a share of stock A, it has a lower bid-ask spread. Investors are limited to trading a single share of stock, so they must choose which asset to trade, or mix between the two. This limitation can be thought of as a capital limit. For a given level of capital, if investors trade the ETF they will receive lower exposure to their specific stock. With this limitation, there are only two cases of equilibria. The first case is a separating equilibrium in which A-informed investors 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 investors will randomize between trading A and trading the ETF. When the bid-ask spread of A and the weight of A in the ETF are sufficiently high, the pooling equilibrium prevails. In the pooling equilibrium, the market maker will have to charge noise traders in the ETF a spread to cover loses from informed trader orders in the ETF. An ETF trade will also cause the market maker to learn about the value of A and update quotes accordingly.

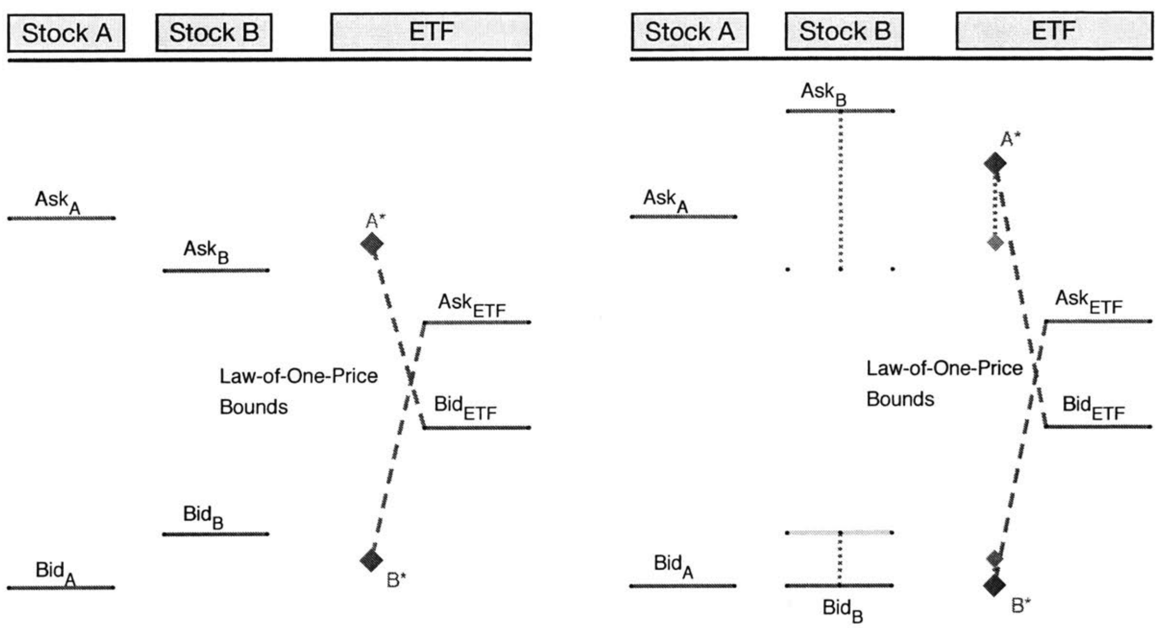
The direct modeling of bid-ask spreads allows this paper to avoid law of one price violations while allowing ETFs and underlying stocks to have different levels of adverse selection (Figure 1). The lack of arbitrage opportunities in the paper may surprise some readers, as the word arbitrage is often used in reference to ETFs. In these industry applications of the term arbitrage, however, there are no true law-of-one-price violations.[1] Having a model with no arbitrage opportunities is

[1]The creation/redemption mechanism is sometimes referred to as an "arbitrage" mechanism. Treating ETF creations/redemptions as law-of-one-price violations is tantamount to referring to mutual fund flows as "arbitrages." After the close of the market, authorized participants (APs) can exchange underlying baskets of securities for ETF

both consistent with the foundations of asset pricing theory, and reflects the market reality for ETFs.[2]

Figure 1. Let A^* and B^* represent the weighted average ask and bid prices of the underlying basket of stocks. No arbitrage gives $Bid_{ETF} < A^*$ and $Ask_{ETF} > B^*$. This set of bounds is depicted on the left.

On the right, an increase in the spread for Stock B leads to wider bounds on the ETF. The existing ETF bid and ask are well within the no-arbitrage bounds, and the quotes could could move up, down, or not at all without triggering any arbitrage opportunities. This paper develops an understanding for how quotes should move together, and thus how the liquidity of each asset is interconnected.



For the full model model, there is also a class of investors with private information about stock B. Both A-informed and B-informed traders have a strategic choice of which asset to trade. There are now two additional equilibria: a partial separating equilibrium and a fully pooling equilibrium.

In the partial separating equilibrium, only one class of informed investors sends mixed orders.

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 are not. Instead, they arise from the technical details of the iNAV calculation. iNAV is usually calculated from last prices of the components, so a deviation from iNAV is typically staleness in prices. 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 will differ from the risk of placing many limit orders in 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, will differ from the market price of the current portfolio. Finally, errors are common in the calculation and reporting of iNAV values Donohue (2012).

[2]KCG analysis on single day of trading for the entire universe of US equity ETFs finds that arbitrage opportunities occur in less than 10% 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)

As an example, investors with information about stock A could trade stock A and the ETF, while investors informed about B only trade stock B. In this setting, the positive spread in the ETF created by the A traders will reduce potential profits that B traders could make in the ETF. This reduction can become so severe that the ETF spread is larger than any profits B traders could make.

This leads to a simple, but novel, theoretical prediction. Traders in the small-weight stocks can be "excluded" from trading the ETF whenever the value of their information is less than cost of the adverse selection they would face from trading the ETF. Small-stock traders have private information which is not incorporated into the ETF price. But they cannot profitably trade the ETF because they face adverse selection from traders with larger, though orthogonal, information. As an example from the XLY Consumer Discretionary ETF, investors in a stock like Gap (0.27% of XLY) who decided to trade XLY would face adverse selection from investors in Amazon (22% of XLY).

In the fully-pooling equilibrium, A-informed and B-informed investors both randomize between trading a single stock and trading the ETF. An ETF trade will cause the market maker to update the quotes for both A and B. The size of the adjustment depends on both the ETF weight and the strength of the market maker's prior. Thus quotes in A and B are both adjusted, but the magnitude of adjustment can differ for the two stocks. As a result, ETF price changes do not cause excessive co-movement of the underlying stocks.

The model gives two easily testable predictions: investors with stock specific information should trade both their stock and the ETF, and this should cause stock-specific adverse selection in the ETF. I test both these hypotheses with high-frequency evidence from NYSE TAQ data from September 1, 2016 to August 14, 2018. I focus on the ten Sector SPRD ETFs from State Street, as well as the 500 underlying securities they hold. The Sector SPDR ETFs have the advantage of being very liquid, fairly concentrated, and representative of a broad set of securities.

To show that investors with stock-specific information also trade ETFs, I look at simultaneous trades between an ETF and its underlying constituents. Under the model, investors should trade both the stock and ETF whenever stock weight in the ETF is high and the stock-specific return or spread is high. Consistent with this prediction, I find that a large stock-specific return or spread predicts a large increase in simultaneous trades with the ETF, and only for the heavily-weighted stocks. This fits the model prediction that heavily-weighted stock investors trade the ETF while lightly-weighted stock investors do not. I am also able to rule out that the alternative story that the simultaneous trades are related to arbitrage activity. 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 result also holds even after controlling for the total number of buy or sell orders in both the stock and ETF.

To show that investors in the ETF are exposed to stock-specific informational asymmetries, I look at the correlation between ETF spreads and underlying constituent spreads. When investors begin to trade the ETF based on their stock-specific information, market makers should charge a spread based on this stock-specific informational asymmetry. As a result, the ETF spread should be correlated with the spread in the individual security. Consistent with the model prediction, I find that a large stock-specific return or spread predicts an increase in the correlation between the stock spread and the ETF spread, and only for the heavily-weighted stocks. Thus part of the spread that noise traders in the ETF pay is due to stock-specific informational asymmetries.

Both these tests are a difference-in-difference. I examine how the stock-ETF interaction changes depending on the stock-specific liquidity, and compare the large-stock/ETF relationship with the small-stock/ETF relationship. As an example, consider the Energies Sector SPDR, XLE. Exxon Mobile (XOM) comprises 23% of the ETF holdings while Cimarex Energy (XEC) comprises just 0.5% of the ETF holdings. The model suggests investors in Exxon Mobile should trade both XLE and XOM when the impact of trades XOM is large, while investors in XEC should only trade XEC regardless of the impact of trades in XEC. Consistent with this prediction, I find that a large return or spread in XOM leads to a large increase in simultaneous trading with XLE, while a large return or spread in XEC has little effect. Similarly, I find a large return or spread in XOM leads to an increase in the correlation of XOM spreads with XLE spreads, while a large return or spread in XEC has little effect.

The rest of the paper is organized as follows. Section II discusses the prior literature. Section III presents the model. Section IV presents empirical evidence on simultaneous trades. Section V presents empirical evidence on spread correlations. Section VI concludes.

II. Literature Review

Several recent papers have created models for price discovery in basket securities, with the earliest literature written with index funds in mind. Gorton and Pennacchi (1991) finds that the different pieces of private information get averaged out in the index fund, and thus liquidity traders can avoid informed traders by trading index funds. Subrahmanyam (1991) takes the distribution of liquidity traders as given, and analyzes how informed traders choose to acquire information in response. As more liquidity traders choose the basket security, informed traders focus on common factors and acquire less security-specific information. More recently, Bhattacharya and O'Hara (2016) considers the information linkage between ETFs and the underlying assets. They construct herding equilibria, where the signal from the ETF overwhelms any signal from the underlying assets, or the signal from underlying assets overwhelms the ETF signal. In all these papers, investors are restricted by assumption to trade only one asset. In my model, informed investors are able to trade any asset, and will sometimes take the opportunity to randomize their orders between assets. In consequence, the noise traders in the ETF are exposed to single-stock adverse selection. ETF trades can have an asymmetric impact on the basket of constituents, with some stocks seeing a large change in quotes while others see a small change.

A second line of related literature deals with arbitrages between assets. Malamud (2016) has a model of risky arbitrage between an ETF and the underlying basket. In my model, the market maker is risk-neutral, but stock-specific adverse selection can spill over when single-stock informed traders begin trading the ETF. The bid-ask spread in my model allows there to be different levels of adverse selection between the ETF and the constituents without law-of-one-price violations.

Finally, the model ties into the literature on the interactions between informed traders. In Caballe and Krishnan (1994), informed traders split their orders across multiple assets to hide their trading intentions. Goldstein, Li, and Yang (2013) creates a model where some informed investors are constrained in respect to which assets they can trade. Cespa and Foucault (2014) models illiquidity spillovers between two assets, like an ETF and the basket, where uncertainty in one market leads to uncertainty in another. In my model, both A and B investors can trade the ETF, but the larger weight A-informed can create a wide ETF spread that prevents B-informed from profitably trading the ETF.

On the empirical side, most previous work has focused on time series relationships. Hasbrouck (2003) 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. Cespa and Foucault (2014) investigates information spillovers between the SPY, E-Mini, and S&P 500 during the flash crash of May 6, 2010. Israeli, Lee, and Sridharan (2017) looks at the level of a company's shares owned by ETF's, and finds that when the level increases, the stock price begins to co-move more with factor news. Hamm (2014) ,examining the relationship between ETF ownership and factor co-movement, finds that while companies with poor quality earnings look more like factor or industry returns when the level of ETF ownership rises, companies with good quality earnings do not show this effect. With trade and inventory data, Pan and Zeng (2016) look at arbitrage in bond ETFs; they find that bond dealers use the ETF to reduce their bond inventories. Finally, Huang, O'Hara, and Zhong (2018) collect evidence that suggests industry ETFs allow investors to hedge risks, and thus increase pricing efficiency for stocks.

In contrast to previous empirical work, my paper will use high-frequency data to understand liquidity sharing between the ETF and the individual ETF components. My simultaneous trade evidence is based on the techniques outlined in Dobrev and Schaumburg (2017), which use trade time-stamps rather than noisy return relationships to identify cross-market activity. The use of high-frequency data will avoid the need for exogenous changes in ETF baskets. Most ETF's are essentially volume-weighted, so inclusion/exclusions from the smallest stocks of the S&P500 or Russell 2000 are the only frequent exogenous changes in ETF baskets.

III. Model

A. Price Discovery in a Single Asset

The model is in the style of Glosten and Milgrom (1985) with two stocks and an ETF. The two stocks will be designated A and B. Both stocks will pay a liquidating dividend from: $\{0, 1\}$. Each stock dividend is independent of the other. The ETF will be designated as (AB), and will contain ϕ shares of A and $(1 - \phi)$ shares of B. Since ϕ can take any value in $[.5, 1]$, A will have a larger weight in the ETF compared to B. This is one source of the asymmetry between ETF stocks, with the other source coming from market maker's prior beliefs.

To keep this section as simple as possible, there will be no price discovery in asset B. This assumption will be lifted in section III.B. For the simplified model of this section, $P(B=1)=\frac{1}{2}$. The market maker will have the observable prior belief that $P(A=1)=\delta$. I will also assume that the market maker is competitive, and thus sets quotes with the intent of making zero expected profits on each trade. Thus for each security, the market maker will set an ask price equal to the expected value of the security conditional on receiving an order to buy, and a bid price equal to the expected value of the security conditional on receiving an order to sell.

The market maker will post limit orders in all three securities. A single trader will be randomly selected to make a trade, and this trader will submit a market order to buy or sell a single share of an asset. The unit mass of traders can be divided into three groups:

- Informed Traders: There will be also be a mass μ_A of traders who are privately informed about A. They know the liquidating dividend with certainty. Each informed trader can trade at most once and for only one share, thus traders have no concern for the price impact of their trades.
- Stock A Noise Traders: There will be a mass of σ_A liquidity traders. They buy or sell stock A with a 50/50 probability.
- ETF Noise Traders: There will be a mass of σ_{ETF} liquidity traders. They will buy or sell the ETF with a 50/50 probability.

The comparative volume of the σ_A noise traders in the single stock versus the σ_{ETF} noise traders in the ETF will be treated as an exogenous parameter. It is easily motivated, however, by stock-specific hedging needs. In choosing to hedge, uninformed traders must balance the value of the hedge against the trading loses from asymmetric information. In both the model and the real world, the ETF will always offer a lower spread than specific stocks. Thus uninformed traders with small endowment shocks or risk aversion will trade the ETF, while uninformed traders with larger endowment shocks or higher risk aversion will choose to trade single stocks.

Each trader will only trade a single share of stock, though randomizing their selection is allowed. The traders who receive a signal about A will face a 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. The single-share limitation of trade can be thought of as a cost of capital or risk-limit for their trading strategy. While investors could trade more aggressively in the ETF to obtain the same single-stock exposure, this would require significantly higher capital and exposure to market risk.

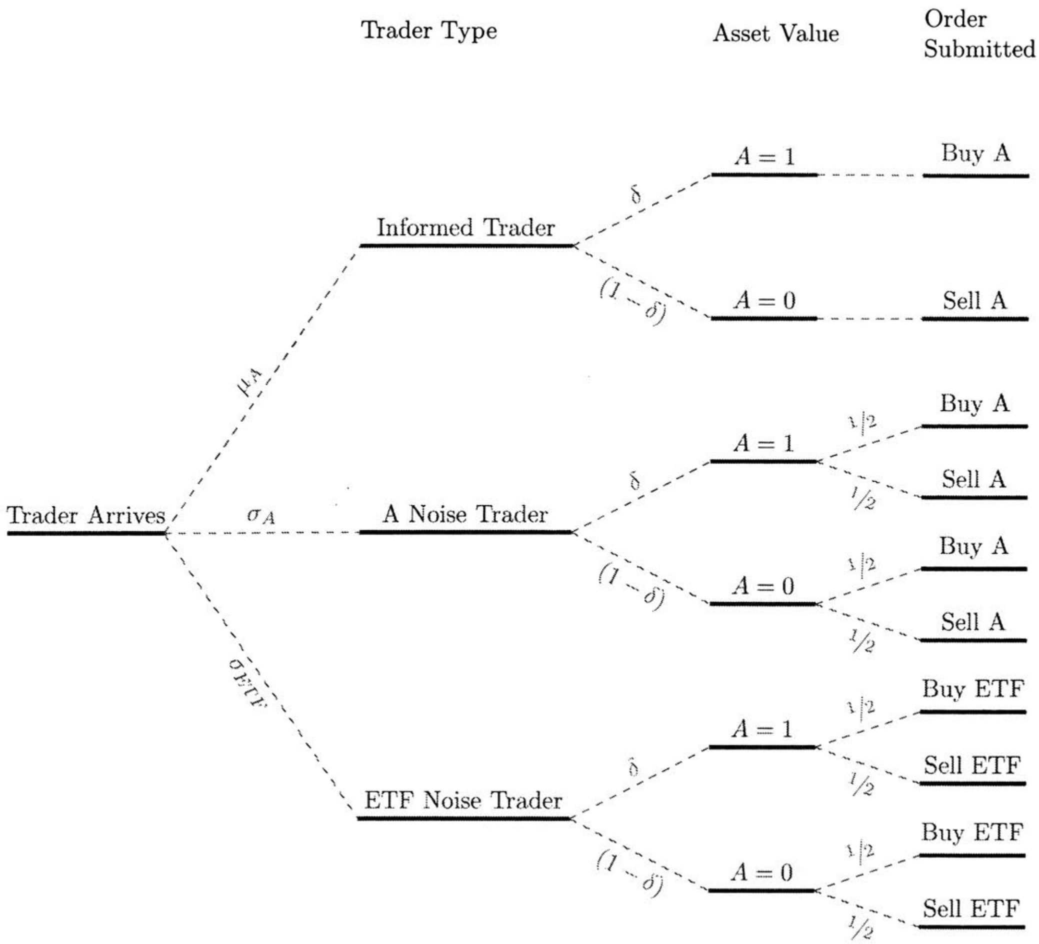
In addition to an ETF weight of each stock, ϕ can be thought of in more general terms as the relevance of the investor's information. When an investor has information about security A, they could also trade a closely related security (AB). While the investor's information is less relevant to the price 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.

Differences in trading strategy can also amplify the size of ϕ. As an example, suppose an investor could invest \$1 in stock A, or invest \$$\gamma$ in the ETF. When the investor invests γ in the ETF, they obtain $\phi\gamma$ of stock A. In this context, $\gamma\phi$ can also be thought of as the substitutability of the single stock and the ETF. A more aggressive trading strategy in the ETF (higher γ) will lead to the ETF to be a closer substitute for the single stock.

Depending on the behavior of informed investors, there are two possible equilibria, outlined in the following propositions. The first is a separating equilibrium, where investors with information about A trade only security A, not the ETF. In this equilibrium, the profits they make from single-stock trading always dominate the profits they could make trading the ETF with no spread. The second is a pooling equilibrium, where investors with information about A mix their orders, randomly sending 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. Note that it is not possible for an equilibrium 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. The sequence of possible trades is reviewed in Figure 2, and key model parameters are reviewed in Table VII in Section A,

PROPOSITION 1: *A separating equilibrium in which informed traders only trade A and do not*

Figure 2. Potential Orders for Separating Equilibrium



trade the ETF, 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 zero spread. For the bid and the ask, these expression are:

$$[\phi \cdot 1 + (1-\phi) \cdot \frac{1}{2} - ask_{(AB)})] \leq 1 - ask_A \tag{1}$$

$$bid_{(AB)} - (1-\phi) \cdot \frac{1}{2} \leq bid_A \tag{2}$$

where the bid and ask prices are the expected value of the stock conditional on an order in the separating equilibrium, and ϕ is the proportion of shares of A in the ETF.

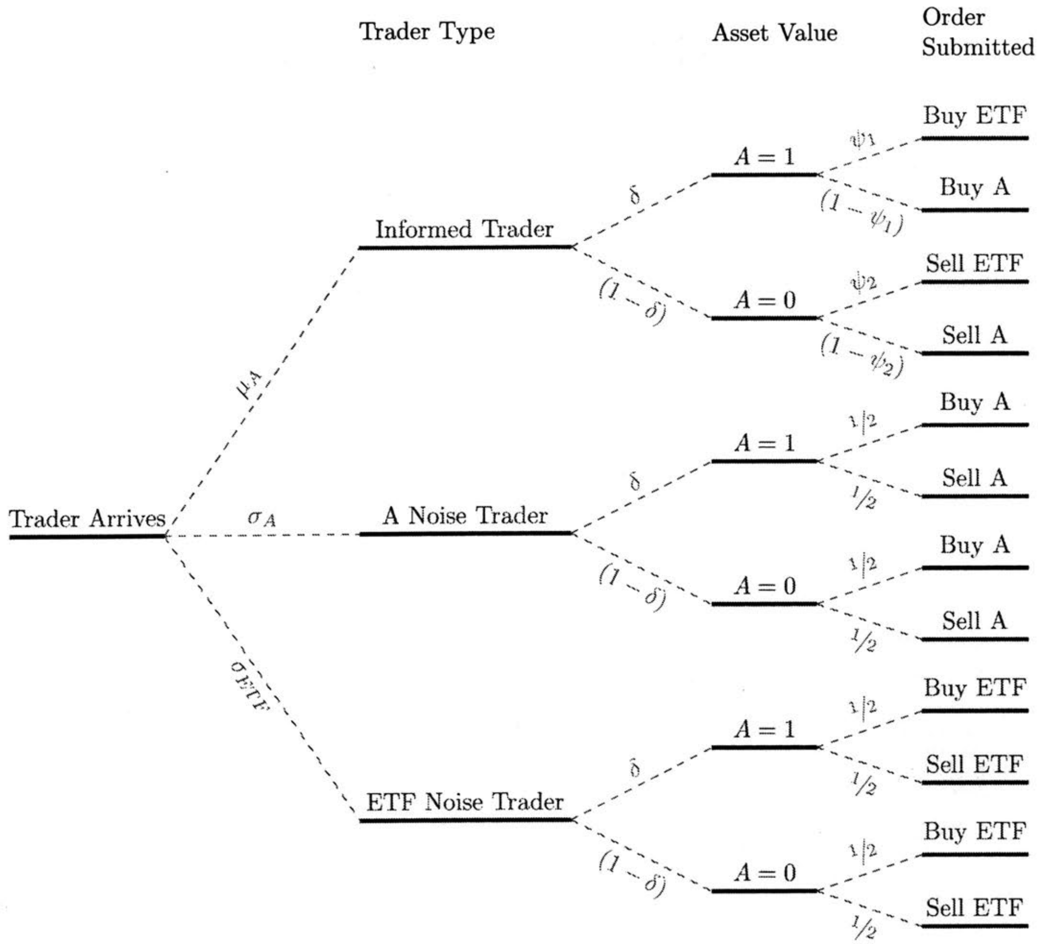
For the market maker to make zero expected profits, each limit order must be the expected value of A conditional on receiving a market order. The asking price is expected value of A conditional on receiving a buy order in A. Since A is worth either 0 or 1, the expected value of A is just the probability that A is equal to 1. A similar logic holds for the bid price. The bid and ask are therefore given by:

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

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 at zero spread 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 investors must randomize between trading the ETF and the single stock. In a pooling equilibrium, informed investors make equal profits trading either the single stock or the ETF. Figure 3 presents the possible trades for the pooling equilibrium.

PROPOSITION 2: *A pooling equilibrium in which informed traders trade both A and the ETF*

Figure 3. Potential Orders for Pooling Equilibrium



exists so long as either of the following conditions hold:

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

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

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

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

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 difference in conditions occurs because the market maker's prior δ may not be $\frac{1}{2}$.

In the pooling equilibrium, an informed trader with a signal about A will randomize 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 mix, the market maker must charge a spread for ETF orders. The ETF noise traders 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 mix, 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 will depend on the probability ψ that he trades the ETF. Shifting more orders to the ETF will increase the ETF spread and decrease 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 volume μ of informed traders and a proportion σ_{ETF} of noise traders in the ETF so 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_{ETF}.

3. The accuracy of the market maker's belief, $-|A - \delta|$.

4. The ETF weighting of the stock, ϕ.

The first three parameters determine the relative sizes of bid-ask spreads. The accuracy of the market maker's belief, $|\delta - A|$, reflects the sensitivity of the market maker's belief to order flow. Suppose, for instance, that the market maker is fairly confident that the true value of A is 1. In the notation of the model, $\delta \approx 1$. The market maker expects informed investors to send buy orders. The asking price, therefore, will be very close to 1, while the bid price will be only slightly less than the estimate of δ. If the true value of the stock is 0, then informed investors will be happy to sell stock A at the price of δ. If the true value of the stock is instead 1, then informed investors face very small profits in the single-stock market. They will instead decide to mix, and send an order to the ETF with a very high probability.

The weight of the stock in an ETF, given by ϕ, determines the potential profit from mixing orders. 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 profits they can make by trading against noise traders in the ETF.

Together, the weighting and the spread size create two separate sources of asymmetry. These two sources of asymmetry lay the groundwork for a difference-in-difference analysis. For the most heavily traded ETFs, stock weights are determined by value weighting. Comparing high-ETF-weight stocks with low-ETF-weight stocks is therefore the same as comparing large market cap stocks with small market cap stocks. As Corollary 1 shows, however, the way in which stocks interact with the ETF also depends on spreads. When large-stock investors face a wide single-stock spread, they find it easy to mix and trade the ETF. When small-stock investors face a wide single-stock spread, they will still find it difficult to mix and trade the ETF. The next section will explore this exclusion of small-stock traders, and the potential for ETF trades to lead to asymmetric revisions in the prices of underlying stocks.

Table I: Equilibrium Spreads

Security	Bid-Ask Quotes
Separating Equilibrium Spreads	
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	$(AB)_{bid} = \phi\delta + (1-\phi)\frac{1}{2}$
	$(AB)_{ask} = \phi\delta + (1-\phi)\frac{1}{2}$

Security	Bid-Ask Quotes
Pooling Equilibrium Spreads	
A	$A_{bid} = \delta\frac{\frac{1}{2}\sigma_A}{(1-\delta)(1-\psi_2)+\frac{1}{2}\sigma_A}$
	$A_{ask} = \delta\frac{\psi_1+\frac{1}{2}\sigma_A}{\delta(1-\psi_1)+\frac{1}{2}\sigma_A}$
ETF	$(AB)_{bid} = \phi\delta\frac{\frac{1}{2}\sigma_{ETF}}{(1-\delta)\psi_2+\frac{1}{2}\sigma_{ETF}} + (1-\phi)\frac{1}{2}$
	$(AB)_{ask} = \phi\delta\frac{\psi_1+\frac{1}{2}\sigma_{ETF}}{\delta\psi_1+\frac{1}{2}\sigma_{ETF}} + (1-\phi)\frac{1}{2}$
Informed Trader	$\psi_1 = \frac{\phi\delta\sigma_{ETF}-\frac{1}{2}(1-\phi)\sigma_A\sigma_{ETF}}{\delta[\sigma_A+\phi\sigma_{ETF}]}$
Mixing Probabilities	$\psi_2 = \frac{\phi(1-\delta)\sigma_{ETF}-\frac{1}{2}(1-\phi)\sigma_A\sigma_{ETF}}{(1-\delta)[\sigma_A+\phi\sigma_{ETF}]}$

B. Price Discovery with Multiple Assets

To develop the full model, I now add price discovery over asset B. Security B can be worth either 0 or 1 and the market maker has a prior belief $P(B = 1) = \beta$. I will also assume that security B is uncorrelated with security A. The unit mass of traders can be divided into five groups:

- Stock A Informed Traders of mass μ_A. They are privately informed about only A.
- Stock B Informed Traders of mass μ_B. They are privately informed about only B.
- Stock A Noise Traders of mass σ_A. They buy or sell stock A with a 50/50 probability.
- Stock B Noise Traders of mass σ_B. They buy or sell stock B with a 50/50 probability.
- ETF Noise Traders of mass σ_{ETF}. They will buy or sell the ETF with a 50/50 probability.

Both classes of informed traders will have a choice to trade one share of any of the securities. Given their stock-specific knowledge, A-informed will consider stock A and the ETF (AB), while B-informed investors will consider 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.

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 mix while investors from the other do not. Without loss of generality, I will examine the case where investors in A trade both A and the ETF (AB), while investors in B only trade stock B.

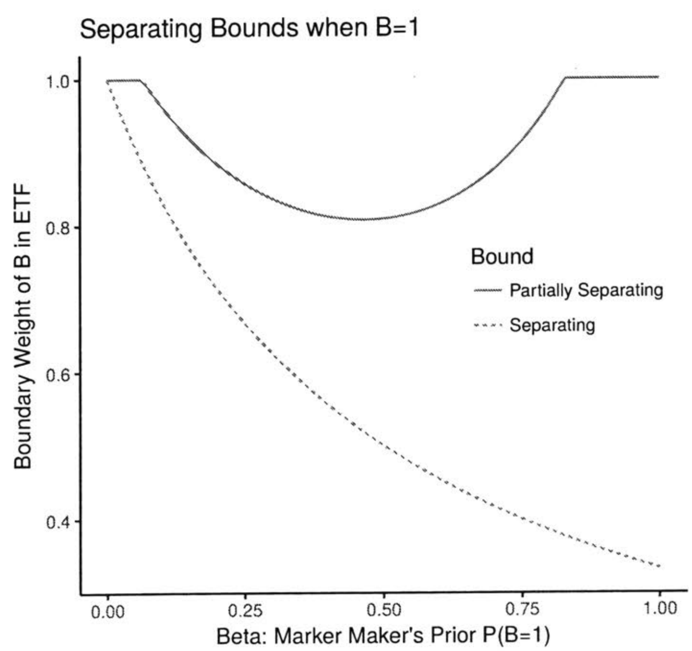
PROPOSITION 3: *partial separating equilibrium. If A traders mix between A and the ETF, B* *traders stay out of the ETF so long as:*

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

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

In the partial separating equilibrium, traders in A mix and behave exactly as they did in the previous section. Traders in B, however, face a different cutoff. Comparing Equation 3 with Equation 5, the cutoff for B traders is much higher than it would be in the absence of the A-informed traders. The reason is that if B-informed traders were to trade the ETF, they would have to pay the adverse selection costs from A-informed traders. The difference in these bounds is illustrated in Figure 4.

Figure 4. If the ETF were available at no spread, B traders will mix whenever the ETF weight of B is above the blue line (Equation 3). Once A traders begin mixing, B traders face adverse selection in the ETF. Their potential profits from ETF trades are reduced, and they only mix when the ETF weight is above the red line (Equation 5).



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, so 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, will set the ETF ask at:

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

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

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

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

The partial separating equilibrium allows disentangling volatility from price impact. In the fully separating equilibrium, only price impact matters, and the cut-off for mixing responds uniformly to the market maker's prior, β. When β is close to the true value of B, the market impact of the trade in B is high. This means the profit from trading the single stock is low, and trading the ETF becomes more appealing. As Figure 4 shows, according as β is closer to B, the ETF weight of B needed for B traders to start mixing is lower.

In the partial separating equilibrium, this is no longer true. Now B traders need to also make enough money from the ETF to overcome the adverse selection from A-informed traders. When β is too close to the true value of B, while the market impact of trades in B is high, the potential profit to be made in the ETF is also small. For B traders to mix, the profits from trading the ETF must be large enough to cover the adverse selection from A traders. They need the true value of β to be sufficiently far from the true value of B. This is akin to B having greater volatility. The B traders will mix only when both the weight and the volatility of B are sufficiently high.

The cutoffs given in Proposition 3 are about the relative profitability between the ETF and stock B. In the model, informed traders can only trade a single asset. In the separating equilibrium, trading only the specific stock earns more profit than randomly trading either the ETF or the stock. It might seem as though the partial separating equilibrium is due to the single-asset trade limitation. This is not the case, however, as the following corollary explains.

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): $$\phi\left(\frac{(1-\delta)\delta\mu_A\psi_2}{(1-\delta)\mu_A\psi_2+\frac{1}{2}\sigma_{ETF}}\right) \geq (1-\phi)\beta$$

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

The adverse selection from A-informed can become so severe that B-informed would loose money trading the ETF. If the ETF was the only asset B-informed could trade, they wouldn't make any trades. Their exclusion from the ETF occurs because investors in A have information that is more important to the ETF price. The 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 can lead to a wide ETF spread from A-informed, and at this spread B-informed would loose money.

While the exclusion of B-informed traders from the ETF is simple, it is novel. B-informed traders have information which is not incorporated into the ETF price. They are risk neutral and have easy access to the noise traders of the ETF, and they have no impatience or concern about the price impact of their trades. They can also be excluded from trading B stock, so that they have no outside option. But they cannot profitably trade the ETF because they face adverse selection from traders with larger, though orthogonal, information.

The exclusion of small-stock traders means that signals are not fully exploitable. Even if a signal predicts an asset return better than the information contained in market prices and investors have infinitely long horizons (so that they will hold the asset until all information is public and reflected in the market price), trading on the signal may not be profitable if the face of adverse selection from other pieces of information. Investors with private information cannot trade every asset correlated with their information; instead, they can only trade assets where the value of their info exceeds the adverse selection from other traders.

The exclusion of small-stock traders has implications beyond ETFs. As an example, consider a stock which has unknown loadings on several different publicly observable factors. Trading the stock based on private knowledge of a factor loading is analogous to trading an ETF based on private knowledge of a single stock. Investors with knowledge of the smaller loadings face adverse selection from investors with knowledge of the larger loadings. Just as in Corollary 2, if the adverse selection becomes too severe, they can be left with no profitable trading opportunities. Even if they are risk-neutral investors with private information that isn't reflected in current prices, there is no way for them to profitably trade.

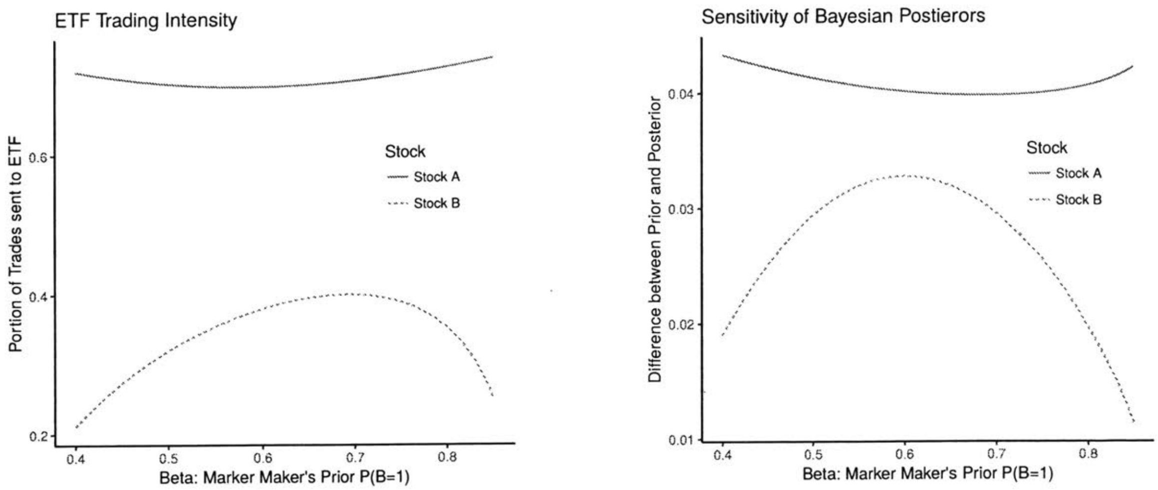
PROPOSITION 4: *If the conditions of Proposition 3 are violated for both securities, then there is a fully pooling equilibrium. Both traders trade the ETF, and following an ETF trade, the market maker has the following Bayesian posteriors:*

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

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

In a fully pooling equilibrium, both informed traders use a mixed strategy of trading the ETF and trading the single stock. Since both investors are mixing, an ETF trade could come from either informed trader. Following an ETF trade, the market maker will change his beliefs about the value of both A and B. The change in beliefs depends on both the weighting and the volatility of the stock in the ETF.

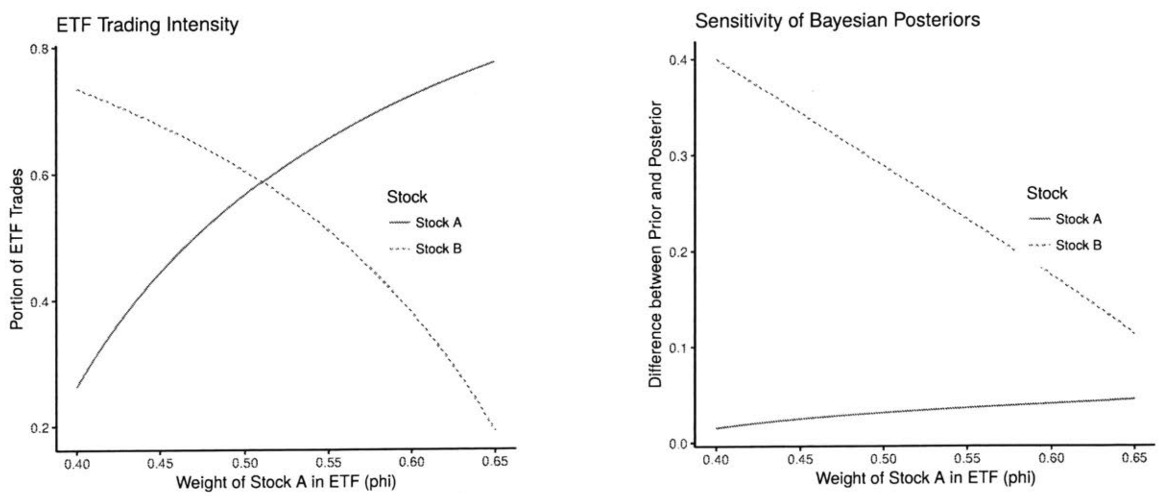
Figure 5 highlights the effect of changes in the market maker's prior. The trading behavior of traders responses in a non-linear way. In Panel 5a, at low levels of β, an increase in β causes B-informed to send more orders to the ETF. This is because a higher β leads to wider quotes in the single-stock market, and the ETF becomes comparatively more appealing. Trading the ETF, however, requires paying the adverse selection costs from A-informed. At very high levels of β, however, the potential profits of B-informed become small relative to this adverse selection cost. The relationship reverses, and an even higher β leads B-informed send fewer of their trades to the ETF.



(a) The market maker's prior β affects profits in both the single stock and the ETF.

(b) The sensitivity of the posterior depends on trading behavior as well as the prior.

Figure 5. Differences in ETF trading and Market Maker's posterior as a function of his prior $P(B=1)$. 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. $\phi = 60$), and the market maker has a prior $\delta = P(A=1) = .6$.



(a) Portion of trades sent to the ETF responds uniformly to changes in ETF weights.

(b) Changes in ETF-weighting have a steeper effect on stocks with a more uncertain prior.

Figure 6. Differences in ETF trading and Market Maker's posterior as a function of the ETF weight of Stock A (ϕ). The securities are symmetric in trader masses: $\mu_a = \mu_b = .15$ and $\sigma_A = \sigma_b = .15$. The priors differ: $P(A=1) = .8$ while $P(B=1) = .75$.

For a standard Bernoulli trial, we would expect the largest change around $\beta = .5$, when the market maker is most uncertain about the value of the security. With the fully pooling equilibrium, however, behavior of the B-informed in trading the ETF also depends on the prior. As Panel 5a shows, trading activity peaks at $\beta \approx 0.7$. The resulting sensitivity of the market maker's prior to an ETF trade, as shown in Panel 5b, peaks at around $\beta = .6$, reflecting a balance of the market maker's initial uncertainty with the trading strategy of informed traders.

Figure 6 highlights the effect of changes in the ETF weight. Higher ETF weight always leads informed traders to send more orders to the ETF (Panel 6a). When the weight is higher, the trader can make more profit and the adverse selection from other securities decreases. For stocks with an uncertain prior, a small change in ETF weight will have a large change in Bayesian sensitivity (Panel 6b, Stock B). For stocks with a more certain prior (Panel 6b, Stock A), small changes in ETF weight will have only small changes in Bayesian sensitivity.

Both the partial separating and the fully pooling equilibrium underscore the importance of volatility in informed investors' decision to trade the ETF. Once the ETF has some informed trading, only investors with information about a sufficiently volatile stock will be able to profit from trading the ETF. The adverse selection will keep out any investors with information about smaller or less volatile stocks.

IV. Empirics

A. Data

This paper analyzes adverse selection spillovers between individual stocks and ETFs. To do this, I examine underlying stocks and the ETF shares of the ten Sector SPRD ETFs from State Street. The Sector SPDRs have the advantage of being very liquid, fairly concentrated, and representative of a broad set of securities. The Sector SPDRs divide the 500 SP500 stocks into ten industry groups: Financials, Energy, Health Care, Consumer Discretionary, Consumer Staples, Industrials, Materials, Real Estate, Technology, and Utilities. Within each ETF, constituents are weighted according to their market cap. As a result, the ETFs have fairly concentrated holdings, as depicted in Figure 7, with the stocks with more than 5% weight comprising between 25% and 50% of each ETF.

Sector SPDRs are extremely liquid. All of them are in the top 100 most heavily traded ETFs, with 6 in the top 25 most traded ETFs. Trading volume 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. From Panel 7b, 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 \$218 million 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.

Microsecond TAQ data was collected on all trades of the ten Sector SPDR ETFs as well as all trades in the stocks that comprise them. The sample period is from September 1, 2016 to August 14, 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. Summary statistics on each stock are presented in Table II.

B. Simultaneous Trades: Basic Setting

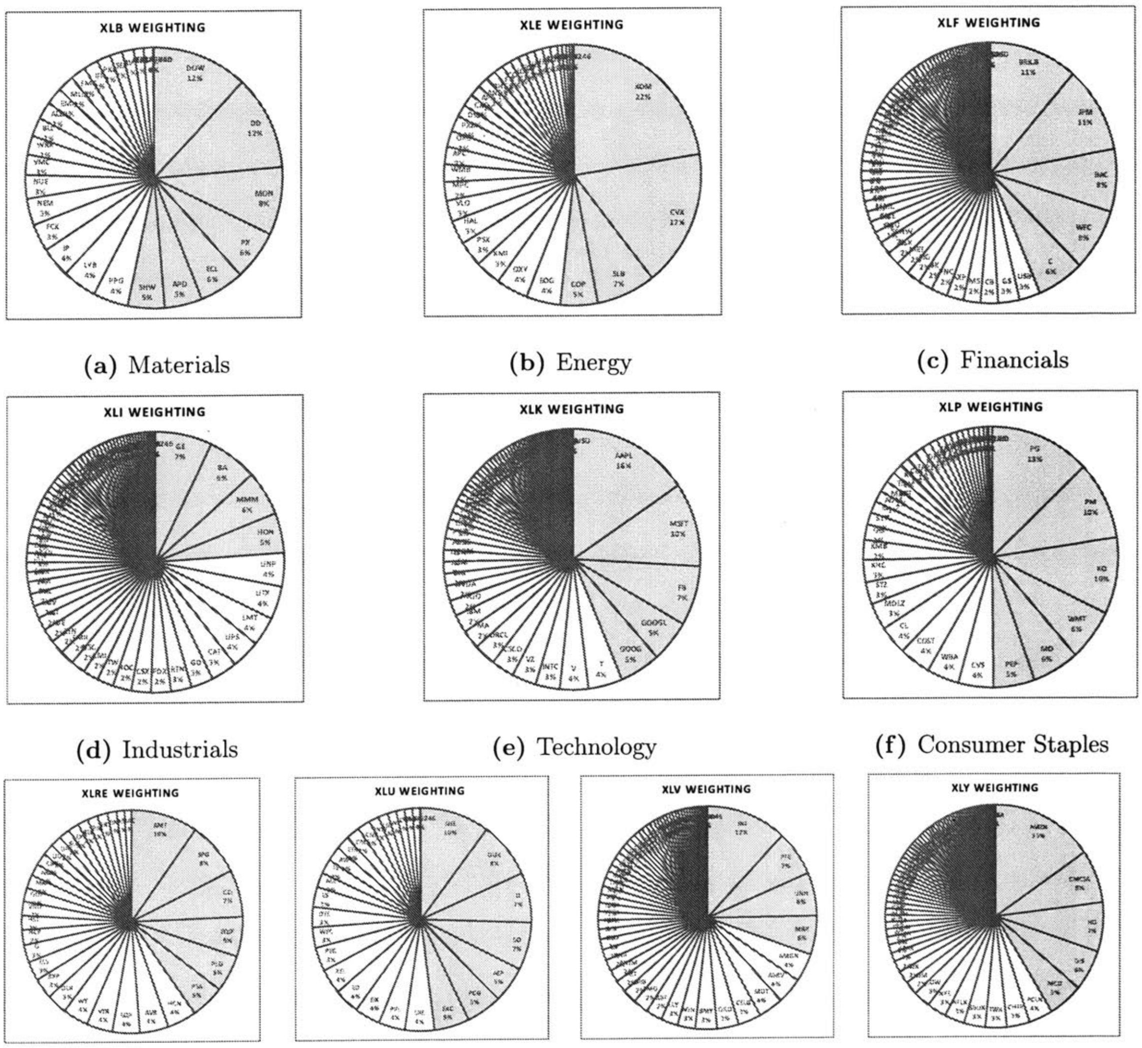
I wish to test the hypothesis that single-stock informed traders will also trade the ETF. Under my theory, investors with stock-specific information mix when both their single-stock market impact is high and their information has sufficient weight in the ETF. This suggests a simple difference-in-difference estimate. When stock-specific traders face a large market impact, they will also trade the ETF only if their stock has a heavy weight in the ETF. Traders in small-weighted stocks will not trade the ETF, even when their stock-specific information is large (Corollary 2).

I will be using anonymous TAQ data, so I can't directly observe the trading behavior of informed participants. I can, however, identify simultaneous trades and conjecture that they belong to the same trader. The 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. In my setting, I will seek to identify trades by the same market participant by looking for simultaneous trades in both a specific stock and the ETF. [3] Trades will be defined as simultaneous if the ETF trade

[3] In Dobrev and Schaumburg (2017), cross market activity counts the portion of times a trade occurs in simultaneously in both market A and market B out of the total number of times a trade occurs in either market. This same measure can be calculated for each offset, t, where A trades t microseconds before B or when A trades t microseconds after B. Formally, cross-market activity, χ, is defined as:

$$\chi_t^{rel} = \frac{\sum_{i=t}^{N-|t|} \mathbf{1}_{\{\text{market A active in period } i\} \cap \{\text{market B active in period } i+t\}}}{\max[\sum_{i=t}^{N-|t|} \mathbf{1}_{\{\text{market A active in period}\}}, \sum_{i=t}^{N-|t|} \mathbf{1}_{\{\text{market B active in period}\}}}$$

In my setting, the maximum in the denominator poses problems. For small stocks, the ETF will always have more trades, while large stocks can sometimes overtake the ETF. To allow a consistent comparison between the ETF-large stock and the ETF-small stock relationships, I avoid normalizing in the definition of cross market activity. Instead, I



(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 7. 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 II: Summary Statistics on Securities

(a) Panel A: Stock Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
Daily Simultaneous Trades	236,670	150.577	227.034	0	6,861
Daily Mean Spread	236,670	0.054	0.100	0.008	3.240
Daily Mean Impact	236,670	0.021	0.028	−0.501	1.049
Daily Mean Realized	236,670	0.004	0.024	−8.670	0.675
ETF Weight	236,670	0.019	0.026	0.001	0.238
Daily ETF Orders	236,670	42,631.290	27,478.750	447	235,645
Daily Stock Orders	236,670	24,111.630	26,808.970	759	1,204,315
Daily Return	236,670	0.145	4.256	−1.000	151.749

(b) Panel B: Correlation between Stock Variables

	Simultaneous Trades	Mean Spread	Mean Impact	Mean Realized	ETF Weight	ETF Orders	Stock Orders	Return
Simultaneous Trades	1	-0.166	-0.152	-0.076	0.426	0.517	0.477	0.021
Mean Spread	-0.166	1	0.862	0.401	0.007	0.011	-0.172	-0.016
Mean Impact	-0.152	0.862	1	0.256	0.003	0.058	-0.144	-0.021
Mean Realized	-0.076	0.401	0.256	1	-0.043	-0.023	-0.093	-0.006
ETF Weight	0.426	0.007	0.003	-0.043	1	-0.014	0.367	-0.022
ETF Orders	0.517	0.011	0.058	-0.023	-0.014	1	0.171	0.034
Stock Orders	0.477	-0.172	-0.144	-0.093	0.367	0.171	1	0.011
Return	0.021	-0.016	-0.021	-0.006	-0.022	0.034	0.011	1

(c) Panel C: ETF Summary Statistics

	ETF	Small Stocks	Large Stocks	Mean ETF Spread	Std Dev. (Spread)	Mean ETF Return (%)	Std Dev. (Return %)	Mean ETF Impact	Std Dev. (Impact)
1	XLV	59	4	0.009	0.0004	0.028	0.708	0.006	0.001
2	XLI	66	6	0.009	0.001	0.027	0.757	0.005	0.001
3	XLY	69	6	0.010	0.001	0.041	0.702	0.006	0.002
4	XLK	69	4	0.009	0.001	0.030	0.828	0.005	0.001
5	XLP	26	4	0.009	0.001	0.010	0.636	0.005	0.001
6	XLU	22	3	0.009	0.0004	0.042	0.777	0.006	0.001
7	XLF	60	4	0.009	0.0004	-0.008	0.848	0.003	0.002
8	XLRE	25	3	0.010	0.002	0.031	0.786	0.002	0.002
9	XLB	19	3	0.009	0.0004	0.007	0.767	0.005	0.001
10	XLE	26	5	0.009	0.001	-0.008	0.932	0.006	0.001

and the single-stock trade occur within 20 microseconds of each other. The lowest latency traders will be colocated with the exchange matching engine. For these traders, sending a message of a trade from the matching engine to the co-located server followed immediately by sending a message from the co-located server back to the matching engine will take more than 20 microseconds. Thus two trades occurring within 20 microseconds of each other cannot be one trade responding to another.

Figure 8. Cross Market Activity in Exxon Mobile Stock on September 1, 2016. The x-axis depicts the offset in microseconds between ETF trades and Exxon trades. The y-axis depicts the number of trades which occur at that exact offset. There is a large amount of cross-market activity between Exxon Mobile and XLE, with 34 trades stamped at exactly the same microsecond (solid red line).There is little to no spike between Exxon Mobile and XLF, an unrelated ETF (dashed blue line). This paper will analyze how this cross-market activity spike changes with day-to-day stock-specific liquidity, and compare the results between stocks with different ETF weights.

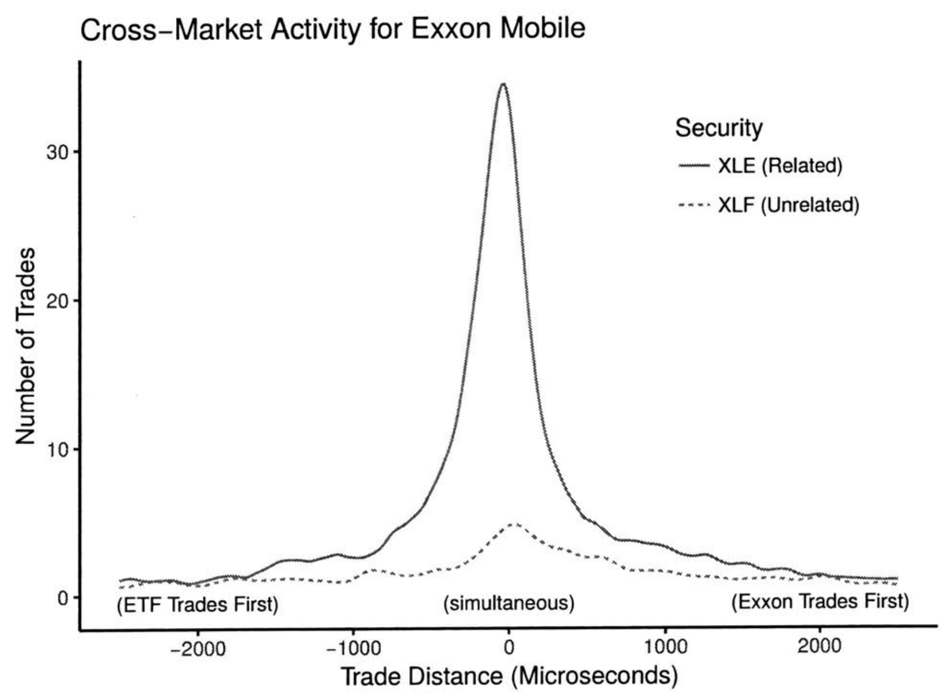


Figure 8 plots an example observation of cross market activity. The spike at exactly 0 microseconds departs form a uniform distribution. These simultaneous trades must have a non-random source, which I assume to be the same investor trading both securities. I will use the level of simultaneous trades as a measure of the level of mixing by stock-specific informed investors. They

control for the total number of stock orders and ETF orders in the daily-level regressions. This has the added benefit of simplifying the interpretation of coefficient estimates.

should mix whenever their stock-specific price impact of trades is high. I will use consider three variables to approximate this: daily stock-specific return, daily average spread, and daily average market impact. This leads to three variations of the same regression:

REGRESSION 1: *For stock i and day t:*

$$Simultaneous\_Trades_{it} = \alpha_0 + \alpha_1 Large\_Return_{it} + \alpha_2 Heavy_i * Large\_Return_{it} + \alpha_3 Controls_{it} \quad (7)$$

$$Simultaneous\_Trades_{it} = \alpha_0 + \alpha_1 Large\_Spread_{it} + \alpha_2 Heavy_i * Large\_Spread_{it} + \alpha_3 Controls_{it} \quad (8)$$

$$Simultaneous\_Trades_{it} = \alpha_0 + \alpha_1 Large\_Impact_{it} + \alpha_2 Heavy_i * Large\_Impact_{it} + \alpha_3 Controls_{it} \quad (9)$$

Heavy is an indicator for stocks which have at least 5% weighting in the ETF. This captures around one-fourth and one-half of the market cap of the Sector SPDRs, and represents five to ten stocks per ETF. Large_Return is an indicator for the days with a stock-specific return among the 5% most positive or 5% most negative of returns for each specific stock over the sample period. Large_Spread is an indicator for the days with the largest 5% of daily mean bid-ask spreads for each specific stock over the sample period. Large_Impact is an indicator for the days in the largest 5% of daily mean market impact for each specific stock over the sample period.

Controls include a fixed effect for each stock, a fixed effect for each day, the total number of stock trades, the total number of ETF trades, and the interaction between the total number of stock and ETF trades, as well as an indicator for either a large ETF return (Equation 7), a large ETF spread (Equation 8), or large ETF impact (Equation 9). An ETF large return indicator takes the value 1 on the 5% most positive and 5% most negative return days for each ETF, while the ETF spread or impact indicators take the value 1 on the 5% of days with the highest spread or market impact for each ETF.

Theory predicts a positive value for α_2: there should be an increase in cross market activity for heavily-weighted stocks on days with high returns. A semi-pooling or fully pooling equilibrium takes place only in the stocks that are sufficiently heavily-weighted or have a sufficiently large informational asymmetry. When the 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 3 and 4). Results for Regression 1 are presented in Table III .

Table III: Estimation of Regression 1

	Dependent variable:		
	simultaneous		
	(1)	(2)	(3)
Large_Return	-3.564^{***}		
	(0.753)		
Large_Spread		-39.030^{***}	
		(1.148)	
Large_Impact			-13.507^{***}
			(1.025)
(α_2) heavy∗Large_Return	81.508^{***}		
	(2.379)		
(α_2) heavy∗Large_Spread		33.198^{***}	
		(3.459)	
(α_2) heavy∗Large_Impact			87.281^{***}
			(3.256)
Observations	239,190	239,190	239,190
R^2	0.822	0.822	0.822
Adjusted R^2	0.821	0.821	0.821
Residual Std. Error (df = 238212)	95.622	95.627	95.679
Note:	*p<0.1; **p<0.05; ***p<0.01		

This table reports estimates of Regression 1, with column (1) showing estimation of Equation 7, colume (2) showing estimation of equation 8 and column (3) showing estimation of Equation 9. The sample is all ten Sector SPDR ETFs and their stock constituents. The frequency of observations is daily. Large Impact is calculated as the median impact from minute-by-minute data. Controls include a fixed effect for each stock, a fixed effect for each day, the total number of stock trades, the total number of ETF trades, the interaction between the total number of stock and ETF trades, and an indicator for either a large ETF return (Equation 7) or large ETF impact (Equation 9). An ETF large return indicator takes the value 1 on the 5% most positive and 5% most negative return days for each ETF, while the ETF spread or impact indicator takes the value 1 on the 5% of days with the highest spread or market impact for each ETF. OLS Standard errors are in parenthesis.

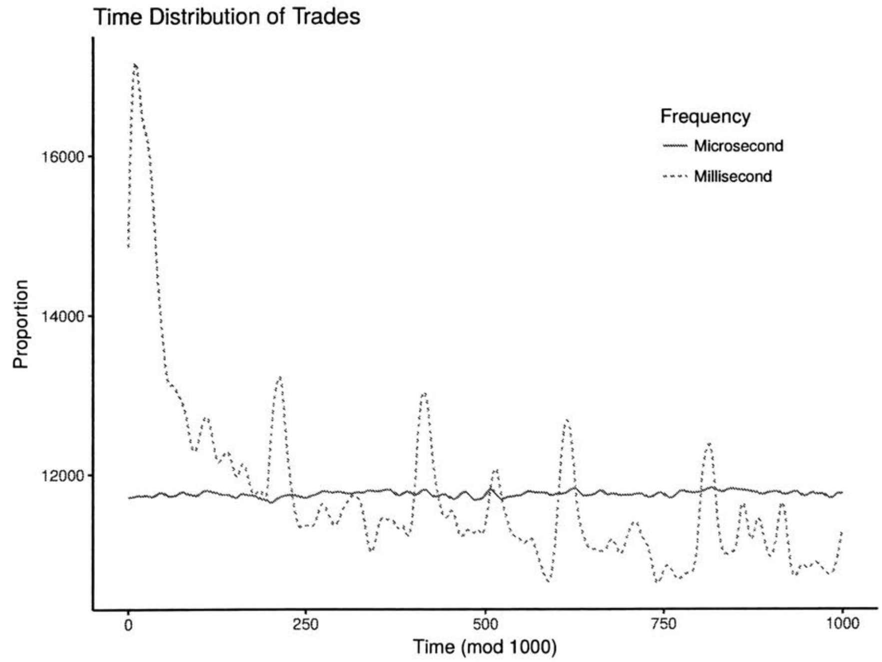
Compared to small stocks, large stocks see an extra 81 simultaneous trades on days when their stock-specific return is high, an extra 33 simultaneous trades when stock-specific spread is high, and an extra 87 simultaneous trades on days when their stock-specific market impact is high. This relationship is both statistically and economically significant. For the set of heavily weighted stocks, the average day has 125 simultaneous trades. Thus on days with high stock-specific informational asymmetries, large stocks see a 25% to 70% increase in simultaneous trades that small stocks do not see. This is consistent with theoretical predictions. From Proposition 4, large stock traders should send more orders to the ETF when their stock-specific price impact is high. From Proposition 3, even when stock-specific volatility or spreads are high, small stock traders will avoid the ETF because they face adverse selection from the large-stock traders.

To check the robustness in the set-up of Regression 1, I replace the indicator variables with the underlying continuous variables. Heavy can be replaced the actual weight of a stock in the ETF. The Large_Return, Large_Spread, or Large_Impact indicators can be replaced with the return, spread, and impact. When I do this, I also replace the ETF indicator control variable with the corresponding continuous value. Results are reported in Table VIII in Section B.A. For returns and price impacts, the results are very similar. The coefficient estimate for α_2 is positive and highly significant. With spreads, the estimates are considerably smaller, and are only significant when the spread is a continuous variable.

An alternative explanation for the source of simultaneous trades might be periodic algorithmic trading. 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, and would seem to be a spike in cross market activity.

At the microsecond level, trades appear to be almost exactly uniform. Figure 9 shows the trade distribution at the millisecond and microsecond level. As a further robustness check, I re-estimate Regression 1 on a restricted subsample. From the sample of all trades, I throw out trades occurring in the first 150 milliseconds and last 50 milliseconds of each second. Of the remaining trades, I throw out and trades occurring within the first 150 microseconds and last 50 microseconds of each millisecond. This leaves trades starting occurring only between $(150, 950)$ of each millisecond to avoid any spikes in activity around 0 milliseconds, and these trades must also fall between

Figure 9. Clock-time periodicities of trades in XLE and the top 10 underlying stocks for September, 2016. I take the timestamp of each trade to the nearest millisecond modulo 1000 to give the blue line, and the timestamp to the nearest microsecond modulo 1000 to give the red line. 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) is almost perfectly uniform.



(150, 950) microseconds to avoid any potential spike in activity around 0 microseconds. Results for this regression are presented in Table IX in Appendix B. For days with large returns or large impact, the results are very similar. For large spread, however, the coefficient is no longer significant. A further robustness check against spurious simultaneous trades will be conducted in the next section, which looks at all possible ETF-stock pairings.

C. Simultaneous Trades: All ETFs

This spike in simultaneous trades should occur between only between an ETF and a specific underlying constituent. It should not occur between a stock and an unrelated ETF. In the second regression, I check the simultaneous trade activity for all possible stock-ETF pairings in my sample.

REGRESSION 2: *For stock i, day t, and ETF j:*

$$Simultaneous\_Trades_{ijt} = \alpha_0 + \alpha_1 Large\_Return_{ijt} + \alpha_2 Heavy_{ijt} * Large\_Return_{ijt} * Proper_{ijt} + \alpha_3 Heavy_{ijt} * Large\_Return_{ijt} + \alpha_4 Controls_{ijt}$$

$$Simultaneous\_Trades_{ijt} = \alpha_0 + \alpha_1 Large\_Spread_{it} + \alpha_2 Heavy_{ijt} * Large\_Spread_{it} * Proper_{ijt} + \alpha_3 Heavy_{ijt} * Large\_Spread_{it} + \alpha_4 Controls_{ijt}$$

$$Simultaneous\_Trades_{ijt} = \alpha_0 + \alpha_1 Large\_Impact_{it} + \alpha_2 Heavy_{ijt} * Large\_Impact_{it} * Proper_{ijt} + \alpha_3 Heavy_{ijt} * Large\_Impact_{it} + \alpha_4 Controls_{ijt}$$

Proper takes the value 1 if the stock i is a member of the ETF j, and takes the value 0 when $j \neq i$. As an example, suppose stock i is Exxon Mobile. Exxon Mobile is in XLE, the Energies Sector SPDR, so the ETF j is XLE, proper takes the value of one. Exxon Mobile is not in XLF, the financials sector SDPR, so if ETF j is XLF, proper takes the value of zero. I consider each stock i from my sample paired against all ten j sector SPDR ETFs.

In this setting, the spike in simultaneous activity should only occur for proper stock-ETF pairings. The coefficient estimate on α_2 should be similar to that in Regression 1. The estimate on α_3, however, should be zero. There should be no spike in simultaneous trades between a stock and an unrelated ETF.

The results of this regression are reported in Table IV. Estimates of α_2 are similar to before, as they should be. The coefficient α_3 for the spike between a stock and an unrelated ETF is very small. While α_3 is statistically significant, in all cases it is around 10% of the parameter estimate of α_2.

D. Signed Trades

One alternative explanation for the spike in cross market activity is that it represents hedging or "arbitrage," where an investor takes opposite positions in the two securities. I can rule this story out by looking at signed trades. 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.

Table IV: Estimation of Regression 2 - All Stock/ETF Combinations

	Dependent variable:		
	simultaneous		
	(1)	(2)	(3)
Proper	81.873***	84.607***	83.510***
	(0.164)	(0.160)	(0.160)
heavy:proper	184.623***	192.931***	189.485***
	(0.556)	(0.542)	(0.542)
Large_Return	−5.488***		
	(0.194)		
Large_Spread		−13.639***	
		(0.277)	
Large_Impact			−9.113***
			(0.261)
Large_Return∗proper	22.145***		
	(0.519)		
Large_Spread∗proper		−11.655***	
		(0.719)	
Large_Impact∗proper			10.945***
			(0.719)
(α_2) heavy∗Large_Return∗proper	93.979***		
	(1.774)		
(α_2) heavy∗Large_Spread∗proper		16.156***	
		(2.446)	
(α_2) heavy∗Large_Impact∗proper			86.894***
			(2.445)
(α_3) heavy∗Large_Return	11.061***		
	(0.609)		
(α_3) heavy∗Large_Spread		1.489*	
		(0.837)	
(α_3) heavy∗Large_Impact			10.546***
			(0.835)
Observations	2,008,091	2,008,091	2,008,091
R^2	0.623	0.622	0.622
Adjusted R^2	0.623	0.622	0.622
Residual Std. Error (df = 2007109)	68.010	68.080	68.086
Note:	*p<0.1; **p<0.05; ***p<0.01		

This table reports estimates of Regression 1, with column (1) showing estimation of Equation 7, column (2) showing estimation of Equation 8, and column (3) showing estimation of Equation 9. The sample is all ten Sector SPDR ETFs and their stock constituents. The frequency of observations is daily. Large Impact is calculated as the median impact from minute-by-minute data. Controls include a fixed effect for each stock, a fixed effect for each day, the total number of stock trades, the total number of ETF trades, the interaction between the total number of stock and ETF trades, and an indicator for either a large ETF return (Equation 7) or large ETF impact (Equation 9). An ETF large return indicator takes the value 1 on the 5% most positive and 5% most negative return days for each ETF, while the ETF impact indicator takes the value 1 on the 5% of days with the highest market impact for each ETF.

To test the hedging theory, I compare the changes in possible simultaneous buys, simultaneous sells, or mixed orders where investors buy or sell the stock and send the opposite order to the ETF. For the stock sign $S_i = \{buy, sell\}$ and ETF sign $Y = \{buy, sell\}$, the four regressions I run follow the form:

REGRESSION 3:

$$\begin{aligned} Simultaneous\_Trades\_Sign\_S\_Y_{it} = \alpha_0 + \alpha_1 Large\_ReturnX_{it} \\ + \alpha_2 Heavy_{it} * Large\_ReturnX_{it} + \alpha_3 Controls_{it} \end{aligned} \tag{10}$$

Trades are signed according to Chakrabarty, Li, Nguyen, and Van Ness (2007), though results with trades signed according to Lee and Ready (1991) are extremely similar. Large will be run in two specifications: $X = LargeN$ for the largest 5% negative returns, and $X = LargeP$ for the largest 5% positive returns. For all regressions, I change the controls from Regression 1 as appropriate. In particular, I have controls for the total number of buy or sell orders, rather than just a single control for orders of any trade sign.

For the largest positive returns, it is buyers of the stock who profit from the days return. Therefore, on those days the informed investors should be buying, and α_2 should be large and positive only for simultaneous buys, and not simultaneous sells or a buy/sell pairing. Similarly for the large negative return days, it is simultaneous sells that should increase, and not simultaneous buys or buy/sell pairs.

Results are presented in Table V. α_2 is much higher for trades of the same sign, where an investor buys both the stock and the ETF at the same time, or sells both the stock and the ETF at the same time. On days when the stock-specific return is highly positive, there is a jump in simultaneous buy orders and only very small jump in simultaneous sell orders or mixed buy/sell orders. This is consistent with the idea that the increase in simultaneous orders is from traders who are informed, as they send liquidity-demanding orders that will profit from the day's trading activity. Similarly, when the stock-specific return is highly negative there is a jump in simultaneous sell orders, a very small jump in buy/sell pairs, and a decrease in simultaneous buys. Again, this is consistent with informed traders taking simultaneous positions in both the stock and the ETF. The positions they take are in the same direction in both securities, and positions which would from the day's price movement.

I also estimate Regression 3 for the indicators on Large_Impact and Large_Spread. Results are presented in Table X in Appendix B. For large price impact, the results are similar. When heavily-weighted stocks have a day when stock-specific market impacts are high, there is a large spike in simultaneous orders of the same sign. On large impact days, heavily-weighted stocks should see an extra 114 simultaneous sell orders and 71 simultaneous buy orders relative to the lightly-weighted stocks. For mixed orders, where investors buy one asset and sell the other, estimates are much lower, at 31 for stock buy/ETF sell and 27 for stock sell/ETF buy. This increase in simultaneous trades is not consistent with hedging or arbitrage activity. For large spreads, the results are less clear. The spike in simultaneous sell orders is much higher than the spike in mixed orders, but there is a decrease in simultaneous buy orders in the heavily-weighted stocks for the large spread days.

Finally, it is worth noting that these results do not rule out that hedging occurs, but only that the simultaneous trades are not hedging. In Huang et al. (2018), for example, investors use the ETF to hedge their industry exposure from single-stock positions. If an investor is going to buy the ETF as well as the single stock, either trade could push up the cost of the other, so executing the two trades simultaneously makes sense. If the investor is going to buy a single stock and use the ETF to hedge, there is no reason to execute the trades within microseconds of each other.

E. Simultaneous Trading Conclusion

Simultaneous trades reflect the idea that single-stock informed traders will also trade the ETF. Consistent with the model prediction, it is investors in large stocks who trade the ETF as well as the single stock. They do so whenever their stock-specific price impact is high and when the value of their information is high.

This increase in simultaneous trades does not show up between a stock and an unrelated ETF, and does not seem to be sensitive to the distribution of trades across time. When trades are signed, the simultaneous orders turn out to be investors buying both securities at once, or selling both securities at once. Thus the results are not consistent with an alternative story of hedging or arbitrage.

Table V : Estimation of Regression 3 - Signed Trades
This table reports estimates of Regression 3. Column (1) shows simultaneous buy orders in the stock and ETF. Column (2) shows a buy order in the stock with a sell order in the ETF. Column (3) shows a sell order in the stock with a buy order in the ETF. Column (4) Shows simultaneous sell orders.
The data set and controls are the same as Regression 1, with the exception that controls on the total number of orders have been changed to the appropriate number of buy or sell orders. Large_Return has been split into two values: LargeP as the 5% most positive returns, and LargeN as the 5% most negative returns.

(a) Panel A: Regression Estimates for Positive Return Days

	Dependent variable:			
Stock Trade Sign:	BUY	BUY	SELL	SELL
ETF Trade Sign:	BUY	SELL	BUY	SELL
	(1)	(2)	(3)	(4)
LargeP	3.563***	−1.343***	−2.773***	−19.349***
	(1.120)	(0.320)	(0.316)	(1.160)
heavy∗largeP	73.215***	16.468***	13.839***	14.202***
	(3.641)	(1.041)	(1.028)	(3.778)
Observations	239,190	239,190	239,190	239,190
R^2	0.780	0.734	0.737	0.770
Adjusted R^2	0.779	0.733	0.735	0.769
Residual Std. Error (df = 238212)	107.192	30.642	30.264	111.201
Note:				*p<0.1; **p<0.05; ***p<0.01

(b) Panel B: Regression Estimates for Negative Return Days

	Dependent variable:			
Stock Trade Sign:	BUY	BUY	SELL	SELL
ETF Trade Sign:	BUY	SELL	BUY	SELL
	(1)	(2)	(3)	(4)
largeN	−18.159***	−2.567***	−0.864***	5.167***
	(1.179)	(0.336)	(0.333)	(1.220)
heavy∗largeN	44.061***	34.676***	29.548***	154.191***
	(3.657)	(1.043)	(1.032)	(3.782)
Observations	239,190	239,190	239,190	239,190
R^2	0.780	0.735	0.737	0.771
Adjusted R^2	0.779	0.734	0.736	0.770
Residual Std. Error (df = 238212)	107.230	30.587	30.224	110.831
Note:				*p<0.1; **p<0.05; ***p<0.01

V. Correlation of Spreads

When investors with stock-specific information trade both the ETF and a single stock, the correlation of bid-ask spreads between the two assets should increase. For example, suppose investors in stock A start trading the ETF as well as stock A. Market makers in the ETF will now face adverse selection form A-informed traders, and must set the ETF spread to cover these losses. Wider spreads in stock A will lead the A-informed to shift a larger portion of their orders to the ETF; this will lead to even wider ETF spreads. As a result, the ETF spread should become more correlated with stock A's spread compared to the other stocks in the ETF. The wider the spread in stock A, the higher the correlation of the ETF spread and stock A's spread.

Motivated by this idea, this section will test the correlation of adverse selection between the ETF and the specific stocks in the ETF. I use the same data as in Section IV. I consider the correlation of three different measures of adverse selection: Realized Spreads, Quoted Spreads, and Price Impact. I calculate each of these measures for each minute of the trading day for each stocks and for each ETFs. For each day and each stock, I calculate the correlation between the stock-specific measure and the ETF measure.

REGRESSION 4:

$$Correlation = \alpha_0 + \alpha_1 Large\_Return_{it} + \alpha_2 Heavy_{it} * Large\_Return_{it} + \alpha_3 Controls_{it} \tag{11}$$

$$Correlation = \alpha_0 + \alpha_1 Large\_Impact_{it} + \alpha_2 Heavy_{it} * Large\_impact_{it} + \alpha_3 Controls_{it} \tag{12}$$

$$Correlation = \alpha_0 + \alpha_1 Large\_Spread_{it} + \alpha_2 Heavy_{it} * Large\_Spread_{it} + \alpha_3 Controls_{it} \tag{13}$$

$$Correlation = \alpha_0 + \alpha_1 Large\_Simultaneous_{it} + \alpha_2 Heavy_{it} * Large\_Simultaneous_{it} + \alpha_3 Controls_{it} \tag{14}$$

As in Section IV, controls include a fixed effect for each stock, a fixed effect for each day, the total number of stock trades, the total number of ETF trades, the interaction between the total number of stock and ETF trades, and either the absolute value of ETF return (Equation 7) or median of 1-minute estimations of ETF spread, price impact, or spread for the day. The first three predictors (Large_Return, Large_Impact, Large_Spread) are the same as those in Regression 1. Based on the results of the previous section, I also consider Large_Simultaneous, which is an indicator which takes the value 1 on days with the largest 5% of simultaneous trades for each stock.

Under the theoretical predictions of the model, α_2 should be positive and significant. Larger spreads should lead to more correlation between the ETF and the specific stock, but only for heavily-weighted stocks. It is only in stocks with a heavy ETF weight that investors will consider trading the ETF with stock-specific knowledge. When investors with stock-specific information start also trading ETFs, the ETF adverse selection should become more correlated with the stock-specific measure of adverse selection.

Results are presented in Table VI. The largest coefficient estimate of α_2 is with price impact (Equation 12). For heavy stocks, the measure of correlation between the ETF and specific stock should increase by .02 on days with a large stock-specific price impact. The mean correlation of quoted spreads for heavy stocks is .3, while the mean correlation of realized spreads and price impacts is .17. Thus the effect is that heavy stocks see an extra 6% increase in correlation with the ETF quoted spread, and a 11% extra increase in correlation with the ETF realized spreads or price impact, relative to the lightly-weighted ETF stocks. Coefficient estimates for the other predictors are smaller, and not significant for two of the settings.

It is important to note that this difference-in-difference setting rules out a common-factor alternative story. Under this story, there could be investors who have private information about a common factor. They may trade both the ETF and the underlying stocks, which would lead to an increase in correlation between ETF and stock spreads. This wouldn't explain, however, the fact that the increase occurs only between the heavily-weighted stocks and the ETF, and not the lightly weighted stocks.

One could argue, however, that factor-informed investors use only the heavily-weighted stocks because they are more liquid. To rule out this story, I conduct an additional test on a restricted set of dates. For each ETF, I calculate the standard deviation of returns among all the underlying ETF constituents. To limit any common adverse selection, I restrict the data to days when the standard deviation of ETF constituents is among the highest 25% of days. On these days, there should be much higher uncertainty about the stock-specific informational asymmetries compared to the common-factor asymmetry. Results from this regression are presented in XII in Appendix B. Coefficient estimates are similar, and in some cases even larger.

These results imply that, contrary to popular belief, investors in ETFs are exposed to stock-

Table VI : Estimation of Regression 4 - Adverse Selection Correlation

This table reports estimates of Regression 4. The sample is all ten Sector SPDR ETFs and their stock constituents. The frequency of observations is daily. Large Impact is calculated as the median impact from minute-by-minute data. Controls include a fixed effect for each stock, a fixed effect for each day, the total number of stock trades, the total number of ETF trades, the interaction between the total number of stock and ETF trades, and either the median minute-by-minute ETF impact or ETF spread or an indicator for large ETF return days.

(a) Panel A: Estimation for Large_Return and Large_Impact

	Dependent variable:					
	Corr_Realized	Corr_Spread	Corr_Impact	Corr_Realized	Corr_Spread	Corr_Impact
	(1)	(2)	(3)	(4)	(5)	(6)
Large_Return	−0.012***	−0.012***	−0.007***			
	(0.001)	(0.002)	(0.001)			
Large_Impact				−0.002**	−0.016***	0.0001
				(0.001)	(0.002)	(0.001)
ETF_Spread	−2.176***	46.716***	−6.060***	−2.078***	46.758***	−6.101***
	(0.295)	(0.658)	(0.290)	(0.294)	(0.659)	(0.290)
heavy* Large_Return	0.004**	0.011**	0.011***			
	(0.002)	(0.005)	(0.002)			
heavy* Large_Impact				0.020***	0.015**	0.019***
				(0.003)	(0.007)	(0.003)
Observations	239,190	239,190	239,190	239,190	239,190	239,190
R^2	0.393	0.557	0.411	0.398	0.557	0.411
Adjusted R^2	0.390	0.555	0.409	0.395	0.555	0.409
Residual Std. Error	0.087	0.194	0.085	0.086	0.194	0.085)
Degrees of Freedom	238214	238211	238211	238211	238211	238211
Note:						*p<0.1; **p<0.05; ***p<0.01

(b) Panel B: Estimation for Large_Spread and Large_Simultaneous

	Dependent variable:					
	Corr_Realized	Corr_Spread	Corr_Impact	Corr_Realized	Corr_Spread	Corr_Impact
	(1)	(2)	(3)	(4)	(5)	(6)
Large_Spread	−0.004***	−0.016***	−0.005***			
	(0.001)	(0.002)	(0.001)			
Large_Simultaneous				0.006***	0.001	0.006***
				(0.001)	(0.002)	(0.001)
ETF_Spread	−1.996***	46.886***	−5.972***	−2.056***	46.649***	−6.059***
	(0.294)	(0.660)	(0.291)	(0.294)	(0.659)	(0.290)
heavy*Large_Spread	0.005*	−0.006	0.007**			
	(0.003)	(0.007)	(0.003)			
heavy*Large_Simultaneous				0.009***	−0.007	0.010***
				(0.002)	(0.004)	(0.002)
Observations	239,190	239,190	239,190	239,190	239,190	239,190
R^2	0.397	0.557	0.411	0.398	0.557	0.411
Adjusted R^2	0.395	0.555	0.409	0.395	0.555	0.409
Residual Std. Error	0.086	0.194	0.085	0.086	0.194	0.085
Degrees of Freedom	238211	238211	238211	238211	38211	238211
Note:						*p<0.1; **p<0.05; ***p<0.01

specific informational asymmetries. When a heavily-weighted stock has a large return day, the ETF spread becomes more correlated with the stock-specific spread. Noise traders in the ETF will therefore pay adverse selection costs based on stock-specific informational asymmetries.

VI. Conclusion

This paper is the first to allow investors with stock-specific information the freedom to trade both ETFs and specific stocks. In an effort to reduce the market impact of their trades, investors will trade both whenever the stock-specific spread and ETF weight are large. On the other hand, traders with information about a small-weight stock are totally excluded from trading the ETF. These small-weight stock traders have information which is not in the ETF price, but the adverse selection they would face in the ETF outweighs any profits they could make.

These results have important implications for ETF pricing. First, noise traders in the ETF will face adverse selection from stock-specific information. Traders in large stocks will continue to have profitable trading opportunities, as they will also be able to earn profits from the ETF noise traders. Second, changes in ETF prices can be distributed asymmetrically to the underlying stocks. Rather than perfect co-movement being the only outcome of ETF price changes, the individual components of the ETF can respond asymmetrically, with some stock prices being very sensitive to the ETF price while other stock prices don't change.

The simultaneous trade evidence supports the theory that investors will sometimes trade both an ETF and a single stock based on stock-specific information. In a difference-in-difference setting, I show that there is a large increase in simultaneous trades on days when stock-specific returns or spreads are high, but only for stocks which have a heavy weight in the ETF. This spike in simultaneous trades does not show up between stocks and unrelated ETFs. When trades are signed, they appear to be simultaneous buy orders in both the stock and the ETF, or simultaneous sell orders. Thus the orders are not consistent with arbitrage.

The correlation between the ETF spread and stock-specific spread provides further support for the idea that noise traders in the ETF face adverse selection from stock-specific information. When a specific stock has a high return or spread, heavily-weighted stocks see measures of adverse selection become more correlated with the ETF spread compared to lightly-weighted stocks. This is consistent with informed trading, and subsequent adverse, in the ETF from large-weight stock-specific information.

The lessons learned from ETFs have implications for broader questions of mapping between information and asset prices. The ETF is an especially easy problem: the weighting of each stock in the ETF is known exactly, and both the ETF and the underlying constituents are continuously traded in a limit order market. If the individual constituents were not traded, the ETF is like a conglomerate firm. If weighting of each constituent is unknown, then the ETF is like a stock with unknown loadings on different factors. The exclusion of small stock traders in the ETF setting thus has a broad set of potential applications. For example, if a stock has a very small and uncertain loading on a factor, the model shows that even risk-neutral traders with knowledge of the exact factor loading and factor return may not have profitable trading opportunities, as they may face adverse selection over unknown and larger aspects of the stock's value.

Appendix A. Proofs

Appendix A. Proof of Proposition 2:

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 cut-off values for a pooling equilibrium to exist.

Table VII: 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 who trade stock A
σ_{ETF}	Fraction of noise traders who trade the ETF
ψ_1	Fraction of informed traders who buy the ETF
ψ_1	Fraction of informed traders who sell the ETF

Appendix B. Proof of Proposition 3: Derivation of The Partial Separating Bound

Let $\sigma_{ETF} = (1 - \mu_A - \mu_B - \sigma_A - \sigma_B)$. 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:

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

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

The ETF bid and ask prices are:

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

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

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.

Appendix C. Proof of Proposition 4: Existence of the Pooling Equilibrium

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

$$\phi(1-\delta_{buy}) + (1-\phi)(\beta - \beta_{buy}) = 1 - \delta\frac{\mu_A(1-\varphi_A)+\frac{1}{2}\sigma_A}{\delta\mu_A(1-\varphi_A)+\frac{1}{2}\sigma_A} \tag{A1}$$

$$\phi(\delta-\delta_{buy}) + (1-\phi)(1 - \beta_{buy}) = 1 - \beta\frac{\mu_B(1-\varphi_B)+\frac{1}{2}\sigma_B}{\beta\mu_B(1-\varphi_B)+\frac{1}{2}\sigma_B} \tag{A2}$$

where

$$\delta_{buy} = \delta\frac{\mu_A\varphi_{A,1} + \beta\mu_B\varphi_{B,1} + \frac{1}{2}\sigma_{ETF}}{\delta\mu_A\varphi_{A,1} + \beta\mu_B\varphi_{B,1} + \frac{1}{2}\sigma_{ETF}}$$

$$\beta_{buy} = \beta\frac{\delta\mu_A\varphi_{A,1} + \mu_B\varphi_{B,1} + \frac{1}{2}\sigma_{ETF}}{\delta\mu_A\varphi_{A,1} + \beta\mu_B\varphi_{B,1} + \frac{1}{2}\sigma_{ETF}}$$

The right side of Equation A1 represents the profits to A-informed investors from from trading the ETF. These profits are decreasing in φ_A. If $\varphi_B = 0$, we would have $\varphi_A = \psi_1$, where ψ_1 is defined in Section A.A. Since ETF profits are decreasing in φ_B, then it must be that $\varphi_A < \psi_1 < 1$.

We also know that if $\varphi_A = \psi_1$, then the violation of Equation 6 in Propsition 3 would imply that B-informed investors have a profitable trading opportunity, and thus $\varphi_B > 0$.

A similar logic applied to Equation A2 gives that $\varphi_B < 1$ and $\varphi_A > 0$.

Now since the right side of Equation A1 is decreasing in φ_A and the left side is increasing, we have a unique φ_A solution. Similarly, Equation A2 gives a unique φ_B solution.

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

Appendix B. Empirics

Appendix A. Robustness Checks to Regression 1

In this section, I present additional results and robustness checks to the empirical analysis.

Table VIII: Robustness Checks of Regression 1- Continuous Variables

	Dependent variable:								
	simultaneous								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Large_Return	−15.892*** (0.872)								
Large_Spread		−34.029*** (1.233)							
Large_Impact			−28.498*** (1.177)						
Absolute_Return				−395.996*** (20.706)			−732.287*** (23.016)		
Median_spread					−198.877*** (5.854)			−195.769*** (6.079)	
Median_Impact						−422.205*** (14.482)			−415.252*** (15.072)
Weight*Large_Return	1,006.282*** (25.533)								
Weight*Large_Spread		6.385 (34.994)							
Weight*Large_Impact			1,028.730*** (34.760)						
heavy*Absolute_Return				2,717.208*** (83.661)					
heavy*Mean_Spread					162.843*** (14.867)				
heavy*Mean_Impact						595.131*** (37.380)			
Weight*Absolute_Return							33,651.070*** (858.220)		
Weight*Mean_Spread								701.964*** (92.609)	
Weight*Mean_Impact									3,043.516*** (260.445)
Observations	239,190	239,190	239,190	239,186	239,190	239,190	239,186	239,190	239,190
R^2	0.822	0.822	0.822	0.822	0.822	0.822	0.822	0.822	0.821
Adjusted R^2	0.821	0.821	0.821	0.821	0.821	0.821	0.822	0.821	0.821
Residual Std. Error	95.557	95.643	95.667	95.612	95.623	95.696	95.501	95.638	95.722
Degrees of Freedom	238212	238212	238212	238208	238212	238212	238208	238212	238212
Note:									*p<0.1; **p<0.05; ***p<0.01

This table reports robustness checks of Regression 1. The sample is all ten Sector SPDR ETFs and their stock constituents. The frequency of observations is daily. Weight is the weight of the stock in the ETF. Price impact is the daily median of the minute-by-minute estimations of the price impact in the stock for a given day. Absolute return is the absolute value of the stock return for a given day. Large_Return is an indicator that takes the value 1 when the stock's daily return is in the largest 5% most positive returns or 5% most negative returns for that stock. Large_Spread and Large_Impact is an indicator for days with daily median spread in the largest 5% for that stock. Controls include a fixed effect for each stock, a fixed effect for each day, the total number of stock trades, the total number of ETF trades, the interaction between the total number of stock and ETF trades, and either the absolute value of ETF return (Equation 7) or median of 1-minute estimations of ETF price impact for the day.

Table IX: Robustness Checks of Regression 1- Time Restricted Sub-Sample

	Dependent variable:		
	simultaneous		
	(1)	(2)	(3)
Large_Return	-12.094^{***}		
	(1.282)		
Large_Spread		-60.912^{***}	
		(1.857)	
Large_Impact			-24.215^{***}
			(1.748)
heavy*Large_Return	57.156^{***}		
	(4.081)		
heavy*Large_Spread		-8.373	
		(5.626)	
heavy*Large_Impact			145.367^{***}
			(5.613)
Observations	236,670	236,670	236,670
R^2	0.822	0.822	0.822
Adjusted R^2	0.821	0.822	0.821
Residual Std. Error (df = 235692)	168.511	168.123	168.288
Note:			*p<0.1; **p<0.05; ***p<0.01

This table reports estimates of Regression 1, with column (1) showing estimation of Equation 7 and column (2) showing estimation of Equation 9. The sample is all ten Sector SPDR ETFs and their stock constituents. The frequency of observations is daily. Large Impact is calculated as the median impact from minute-by-minute data. Controls include a fixed effect for each stock, a fixed effect for each day, the total number of stock trades, the total number of ETF trades, the interaction between the total number of stock and ETF trades, and an indicator for either a large ETF return (Equation 7) or large ETF impact (Equation 9). An ETF large return indicator takes the value 1 on the 5% most positive and 5% most negative return days for each ETF, while the ETF impact indicator takes the value 1 on the 5% of days with the highest market impact for each ETF.

Estimation of Regression 3- Signed Trades

Table X This table reports estimates of Regression 3. Column (1) shows simultaneous buy orders in the stock and ETF. Column (2) shows a buy order in the stock with a sell order in the ETF. Column (3) shows a sell order in the stock with a buy order in the ETF. Column (4) Shows simultaneous sell orders in the stock and the ETF.
The data set and controls are the same 1, with the exception that controls on the total number of orders have been changed to the appropriate number of buy or sell orders.

(a) Panel A: Regression Estimates with Large Market Impact

	Dependent variable:			
Stock Trade Sign:	BUY	BUY	SELL	SELL
ETF Trade Sign:	BUY	SELL	BUY	SELL
	(1)	(2)	(3)	(4)
Large_Impact	−12.761***	−5.217***	−4.925***	−14.738***
	(1.128)	(0.322)	(0.318)	(1.168)
heavy∗Large_Impact	71.781***	31.111***	27.870***	114.195***
	(3.643)	(1.040)	(1.028)	(3.775)
Observations	239,190	239,190	239,190	239,190
R^2	0.780	0.735	0.737	0.770
Adjusted R^2	0.779	0.734	0.736	0.769
Residual Std. Error (df = 238213)	107.212	30.603	30.227	111.055
Note:				*p<0.1; **p<0.05; ***p<0.01

(b) Panel B: Regression Estimates With Large Spreads

	Dependent variable:			
Stock Trade Sign:	BUY	BUY	SELL	SELL
ETF Trade Sign:	BUY	SELL	BUY	SELL
	(1)	(2)	(3)	(4)
Large_Spread	−27.451***	−7.745***	−7.800***	−26.709***
	(1.202)	(0.343)	(0.339)	(1.246)
heavy∗Large_Spread	−21.180***	17.112***	14.080***	44.735***
	(3.653)	(1.043)	(1.032)	(3.790)
Observations	239,190	239,190	239,190	239,190
R^2	0.780	0.734	0.737	0.770
Adjusted R^2	0.779	0.733	0.736	0.769
Residual Std. Error (df = 238213)	107.157	30.625	30.241	111.156
Note:				*p<0.1; **p<0.05; ***p<0.01

Table XII : Estimation of Regression 4 - Adverse Selection Correlation
This table reports estimates of Regression 4. In this table, data has been restricted. For each ETF, I calculate the standard deviation of returns among all the underlying ETF constituents. To limit any common adverse selection, I restrict the data to days when the standard deviation of ETF constituents is among the highest 25% of days.Controls are the same 1, with the addition of a control for the daily mean ETF spread.

(a) Panel A: Estimation for Large_Return and Large_Impact

	Dependent variable:					
	Corr_Realized	Corr_Spread	Corr_Impact	Corr_Realized	Corr_Spread	Corr_Impact
	(1)	(2)	(3)	(4)	(5)	(6)
Large_Return	-0.012^{***}	-0.018^{***}	-0.010^{***}			
	(0.001)	(0.002)	(0.001)			
LimpactP				-0.008^{***}	-0.026^{***}	-0.005^{***}
				(0.001)	(0.003)	(0.001)
ETF_Spread	-1.042	43.341^{***}	-2.727^{**}	-1.061	43.558^{***}	-2.774^{**}
	(1.212)	(2.704)	(1.205)	(1.213)	(2.704)	(1.206)
heavy∗Large_Return	0.014^{***}	0.012	0.013^{***}			
	(0.004)	(0.008)	(0.004)			
heavy∗Large_Impact				0.021^{***}	0.021^{**}	0.017^{***}
				(0.005)	(0.011)	(0.005)
Observations	59,738	59,738	59,738	59,738	59,738	59,738
R^2	0.337	0.602	0.340	0.336	0.602	0.339
Adjusted R^2	0.327	0.596	0.330	0.326	0.596	0.329
Residual Std. Error (df = 58843)	0.082	0.183	0.082	0.082	0.183	0.082
Note:						*p<0.1; **p<0.05; ***p<0.01

(b) Panel B: Estimation for Large_Spread and Large_Simultaneous

	Dependent variable:					
	Corr_Realized	Corr_Spread	Corr_Impact	Corr_Realized	Corr_Spread	Corr_Impact
	(1)	(2)	(3)	(4)	(5)	(6)
Large_Spread	-0.005^{***}	-0.038^{***}	-0.005^{***}			
	(0.002)	(0.004)	(0.002)			
Large_Simultaneous				0.008^{***}	0.002	0.009^{***}
				(0.001)	(0.003)	(0.001)
ETF_Spread	-1.112	43.377^{***}	-2.786^{**}	-1.055	43.172^{***}	-2.742^{**}
	(1.213)	(2.702)	(1.206)	(1.213)	(2.705)	(1.206)
heavy∗Large_Spread	0.014^{***}	0.026^{**}	0.013^{**}			
	(0.005)	(0.012)	(0.005)			
heavy∗Large_Simultaneous				0.012^{***}	-0.004	0.010^{***}
				(0.004)	(0.008)	(0.004)
Observations	59,738	59,738	59,738	59,738	59,738	59,738
R^2	0.336	0.603	0.339	0.336	0.602	0.339
Adjusted R^2	0.326	0.597	0.329	0.326	0.596	0.329
Residual Std. Error (df = 58843)	0.082	0.183	0.082	0.082	0.184	0.082
Note:						*p<0.1; **p<0.05; ***p<0.01

REFERENCES

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

Caballe, Jordi, and Murugappa Krishnan, 1994, Imperfect competition in a multi-security market with risk neutrality, *Econometrica: Journal of the Econometric Society* 695–704.

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.

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).

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.

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.

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

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

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.

Huang, Shiyang, Maureen O'Hara, and Zhuo Zhong, 2018, Innovation and informed trading: Evidence from industry etfs .

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.

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

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

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.

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

Wigglesworth, Robin, 2017, Etfs are eating the us stock market, https://www.ft.com/content/6dabad28-e19c-11e6-9645-c9357a75844a, Accessed: 2017-12-01.