The Sound of Many Funds Rebalancing

Alex Chinco† and Vyacheslav Fos‡

December 3, 2016

Abstract

This paper proposes that long rebalancing cascades (stock A’s price jumps, which causes a fund to buy stock A and sell stock B, which causes a second fund to sell stock B and buy stock C, which causes...) generate noise in financial markets. We show theoretically that, when funds follow many different threshold-based trading strategies, a change in stock A’s fundamentals can trigger long rebalancing cascades that eventually affect the demand for unrelated stocks on the other side of the market, like stock Z. And, we prove that in a large market it’s computationally infeasible to predict whether long rebalancing cascades will result in buy or sell orders for stock Z. The best you can do is compute stock Z’s susceptibility to these erratic, non-fundamental demand shocks. We use data on exchange-traded funds (ETFs) to empirically test this proposal. We find that, when stock A realizes a price shock, stock Z sees a larger increase in ETF-rebalancing volume at times when it’s held by many different ETFs and thus more susceptible to long ETF-rebalancing cascades. And, we document that informed traders in stock Z’s trade more aggressively at precisely these moments, suggesting that market makers view the demand coming from long ETF-rebalancing cascades as noise.

JEL Classification. G02, G12, G14

Keywords. Noise, Exchange-Traded Funds, Index-Linked Investing, Excess Trading Volume, Cascades and Contagion

---

*We thank Nick Barberis, Zahi Ben-David, James Choi, Tarun Chordia, Xavier Gabaix, Itay Goldstein, Tim Johnson, Charles Jones, Alan Moreira, Neil Pearson, Jeff Pontiff, Alexei Tchistyi, and Scott Weisbenner as well as seminar participants at the University of Illinois, Boston College, Yale, CalTech, and the FINRA Market-Structure Conference for extremely helpful comments and suggestions.


†University of Illinois at Urbana-Champaign, College of Business; alexchinco@gmail.com.

‡Boston College, Carroll School of Management; fos@bc.edu.
1 Introduction

Imagine you’re a trader who’s just discovered stock Z is under-priced. In a market without noise, there’s no way for you to take advantage of your discovery. The moment you try to buy a share, market makers will immediately realize that you must have uncovered some good news, so you won’t find any sellers (Milgrom and Stokey, 1982). Noise pulls the rug out from under this no-trade theorem. In a market with noise, some traders are always trading stock Z for erratic, non-fundamental reasons. So, when you try to buy a share, it won’t raise any eyebrows. Market makers won’t immediately realize that you must have uncovered good news because your buy order could just be some more random noise. The existence of this plausible cover story allows you to trade on and profit from your discovery. In other words, noise provides the alibi that “makes financial markets possible” (Black, 1985).

But, where does this noise come from? Who generates it? And, what are these traders’ erratic, non-fundamental reasons for trading?

The standard answer to these questions is that noise comes from individual investors, and individual investors’ trading looks erratic and unrelated to fundamentals because they are bad traders. There are good reasons why this is the standard answer. Barberis and Thaler (2003) catalogue a litany of behavioral biases that individual investors suffer from, and Barber and Odean (2000) document that individual investors trade too often. So, there’s ample evidence that individual investors can generate noise. But, are they the only source? It seems unlikely. After all, the fraction of U.S. equity owned by individual investors has fallen by more than 50% since 1980, and we haven’t seen a corresponding drop in trading volumes (Stambaugh, 2014).

This paper proposes another noise-generating mechanism: long rebalancing cascades (stock A’s price jumps, which causes a fund to buy stock A and sell stock B, which causes a second fund to sell stock B and buy stock C, which causes…). We show theoretically that, when funds follow many different threshold-based trading strategies, a change in stock A’s fundamentals can trigger a long rebalancing cascade that eventually affects the demand for unrelated stocks on the other side of the market, like stock Z. And, we prove that, in a large market, it’s computationally infeasible to predict whether long rebalancing cascades will result in buy or sell orders for stock Z. All you can do is compute stock Z’s susceptibility to this erratic, non-fundamental demand.

Then, we use data on exchange-traded funds (ETFs) to empirically verify this proposal. We find that, when stock A realizes a price shock, stock Z only sees increased ETF-rebalancing volume at times when it’s held by many different ETFs and thus more susceptible to long ETF-rebalancing cascades. And, we document that stock Z’s informed
traders trade more aggressively at precisely these same moments, suggesting that market makers view the erratic, non-fundamental demand coming from long ETF-rebalancing cascades as noise.

_Economic Mechanism._ Threshold-based trading strategies are incredibly common. Hedge funds trade on momentum, buying the 30% of stocks with the highest past return and selling the 30% of stocks with the lowest (Asness et al., 2013). This is a threshold-based trading strategy because an arbitrarily small difference in returns can move a stock from the 31st to the 30th percentile of past returns. The PowerShares S&P 500 Low-Volatility ETF is an exchange-traded fund that holds the quintile of S&P 500 stocks with the lowest volatility. An arbitrarily small difference in volatilities can move a stock from having the 101st to having the 100th lowest volatility. And, pension funds are often required to invest 10% of their assets in alternative strategies. A single bad real-estate investment can leave a pension fund with less than 10% of their assets invested in alternative strategies.

When funds follow many different threshold-based strategies, we show theoretically that a small change in one asset’s fundamentals can trigger a long rebalancing cascade, which can affect the demand for unrelated stocks on the other side of the market. For example, a single bad real-estate investment can force a pension fund to rebalance its portfolio and liquidate some of its equity holdings in NextEra Energy Inc. This sale can generate additional volatility in NextEra Energy’s returns, pushing the stock out of the quintile of S&P 500 stocks with the lowest volatility and forcing the PowerShares S&P 500 Low-Volatility ETF to replace NextEra Energy with Duke Energy Corp. And, this purchase of Duke Energy can increase its returns by just enough so that it’s among the 30% of stocks with the highest past returns.

But, this single shock can trigger many complicated cascades that branch and reconnect with one another. The 3-stock cascade above is only one possible way that the effects of the bad real-estate investment can wind their way through the market and alter the demand for Duke Energy. We show that, in a large market, it’s computationally infeasible to compute the net effect of these many cascades on the demand for Duke Energy. The best you can do is compute how susceptible Duke Energy is to the erratic, non-fundamental demand coming from long rebalancing cascades. The aggregate demand coming from long rebalancing cascades looks like random noise even though each individual rebalancing decision in every one of the cascades is perfectly predictable on its own. In truth, the idea that computational difficulty can generate randomness has a long history in economics and finance. Keynes (1921) pointed out that, although the total population of France isn’t random, whether or not this total is an even or odd number at any given instant may as well be.

_Empirical Analysis._ Having seen how long rebalancing cascades can generate noise in
financial markets, we next use data on ETFs to empirically test this idea. The ETF market is an ideal laboratory for our study for two reasons. First, ETFs track many different benchmark indexes, ranging from large-cap to sector-specific to low-vol and every combination in between. Second, these strategies are often threshold-based. A small change in the volatility of NextEra Energy can push it out of a large-cap, low-vol, energy index.

Here’s how our empirical analysis works. Long rebalancing cascades need to start with some sort of shock—something has to knock over the first domino. We study long ETF-rebalancing cascades that start with M&A announcements, and we use “stock A” as the name of the M&A target where the cascade originates. Then, we look at the trading activity of unrelated stocks on the other side of the market in the days after this announcement. We use “stock Z” as the name of an unrelated stock on the other side of the market from stock A. Stock Z can be any stock that isn’t in the same industry as stock A and isn’t held by any of the same ETFs that also hold stock A.

We present two main empirical results. First, we show that, when stock Z is held by many different ETFs and thus more susceptible to long ETF-rebalancing cascades, ETF rebalancing for this stock increases by 14% after the M&A announcement about stock A. This relationship between ETF-rebalancing volume on one side of the market and shocks to unrelated stocks on the other side of the market wouldn’t exist in a world without long ETF-rebalancing cascades. Because we use stock fixed effects, our results can’t be explained by differences in ETF trading across stocks. We’re looking at the percentage increase in ETF-rebalancing volume for the same stock at times when that stock just happens to be held by a few more ETFs. We also make use of a key institutional detail: ETFs that rebalance daily tend to do their rebalancing in the final 20 minutes of the trading day. And, our results are strongest following M&A announcements where the originating stock, stock A, realizes a spike in end-of-day volume on the day of its announcement.

Second, we give evidence that market makers view the unpredictable demand coming from these long ETF-rebalancing cascades as noise. We follow Collin-Dufresne and Fos (2015) and use Rule 13(d) of the 1934 Securities Exchange Act to identify demand from informed traders. This rule says that shareholders have to let the Securities and Exchange Commission (SEC) know within 60 days whenever they acquire 5% or more of a publicly traded company that they have a voting interest in. Consistent with the idea that market makers view the demand from long ETF-rebalancing cascades as noise, we find that stock Z’s informed traders trade more aggressively at precisely those times when stock Z realizes erratic, non-fundamental demand from long ETF-rebalancing cascades starting with the shock to stock A.
1.1 Related Literature

This paper borrows from and brings together several strands of literature.

*Exchange-Traded Funds.* The ETFs represent an enormous market, and other researchers are actively studying ETFs. But, this existing research tends to look at how ETFs affect prices. For example, in an innovative working paper, Ben-David et al. (2016) show that ETFs increase the return volatility of the stocks they hold. And, Tuzun (2013) suggests that leveraged ETFs are generating crash risk. In this paper, by contrast, we study ETFs’ affect on demand—that is, the way that long ETF-rebalancing cascades generate noise. So, we view these other findings as complementary to our own findings.

*Index-Linked Investing.* There are many excellent papers on the economic consequences of index-linked investing (Wurgler, 2010). To give one example, Barberis et al. (2005) show that a stock’s beta with the S&P 500 jumps sharply after it gets added to the index. To give another, Greenwood and Thesmar (2011) and Vayanos and Woolley (2013) give evidence that stocks which are held by similar funds tend to have more similar returns than expected. In contrast to these papers, we focus on the unpredictable effects of index-linked investing rather than the predictable ways that individual investors do it badly.

*Excess Trading Volume.* Researchers have given several explanations for enormous trading volumes that we observe in modern financial markets. The most common explanations center on disagreement between traders and the mis-interpretation of market signals (Harrison and Kreps, 1978; Hong and Stein, 2007; Scheinkman and Xiong, 2003). Other explanations build on wavering beliefs Barberis et al. (2016). This paper shows how excess trading volumes can arise from the interaction of simple trading rules.

*Cascades and Contagion.* Finally, threshold-effect models have been around in the sociology literature since the mid-1970s (Granovetter, 1978). More recently, Watts (2002) showed how cascades can form in random networks. Papers in the finance and economics literature have also looked at cascades and herding dating back to at least Froot et al. (1992). But, while similar in spirit, our paper differs from this work in an important way. These papers study how cascades amplify shocks; whereas, we study how cascades transmit shocks to unrelated assets in unpredictable ways.

2 Economic Mechanism

Trading strategies don’t exist in a vacuum. For example, Khandani and Lo (2007) point out that “during the week of August 6, 2007, a number of [quantitative hedge funds] experienced
\[ \lambda = 0.80 \quad 0.90 \quad 1.00 \quad 1.10 \quad 1.20 \]

**Figure 1:** Examples of network structure for \( \lambda \in \{0.8, 0.9, 1.0, 1.1, 1.2\} \) and \( S = 100 \). Each top panel shows a single random-network realization for the given value of \( \lambda \). Nodes are stocks \( s = 1, 2, \ldots, 100 \). Edges are the rebalancing rules of different funds. If stocks \( s \) and \( s' \) share an edge, then a shock to stock \( s \)'s fundamentals will cause some fund to rebalance and trade stock \( s' \) instead. Each bottom panel shows both the theoretical distribution (solid line, Lemma 2.1) and the sample distribution (dashed line) of the number of neighbors, \( N_s \), for the given value of \( \lambda \). Reads: “The market is more connected when \( \lambda \) is larger.”

Unprecedented losses. [These] initial losses [were] due to the forced liquidation of one or more large equity market-neutral portfolios. [And, the] subsequent price impact...caused other similarly constructed portfolios [and] led to further losses [and] more deleveraging and so on.” In this section, we write down a model to better understand both when long rebalancing cascades are likely to occur and whether the net effects of these long rebalancing cascades will be easy to predict.

### 2.1 Market Structure

Consider a market with \( S \) stocks, \( s = 1, 2, \ldots, S \), where \( S \) is big. In the analysis below, we will often look at the limiting case where \( S \to \infty \).

**Network of Stocks.** If a change in the fundamentals of stock \( s' \) will cause a fund to rebalance and buy stock \( s \) instead, then we say that stock \( s' \) and stock \( s \) are neighbors. They are connected by a fund’s rebalancing rule. They share an edge. Suppose that two randomly selected stocks are neighbors with probability \( \lambda / S \) where \( 0 < \lambda = O[\log(S)] \). Let \( \mathcal{N}_s \) denote the set of stocks neighboring stock \( s \), and let \( N_s \overset{\text{def}}{=} |\mathcal{N}_s| \) denote the number of neighbors that stock \( s \) has. Lemma 2.1 below characterizes the distribution of neighbors for
each stock as $S \to \infty$.

**Lemma 2.1 (Number of Neighbors).** The number of neighbors for each stock as $S \to \infty$ is Poisson distributed,

$$N_s \sim \text{Pois}(\lambda),$$

with $\lambda = E[N_s] = \text{Var}[N_s]$.

Thus, the parameter $\lambda$ captures how connected stocks are. If $\lambda \approx 0$, then the market is fragmented. A change in one stock’s fundamentals is unlikely to cause any funds to rebalance and buy another stock instead. By contrast, if $\lambda \gg 0$, then the market is densely connected. A change in one stock’s fundamentals will cause many funds to rebalance and trade many other stocks. Figure 1 gives some examples of this network structure for different levels of connectivity when there are $S = 10^2$ stocks.

**Rebalancing Cascades.** If stock $s_A$’s price jumps, then this price change can cause a trading strategy to buy stock $s_A$ and sell stock $s_B$. And, the impact of this sell order on stock $s_B$’s price can cause a second strategy to sell stock $s_B$ as well and buy stock $s_C$ instead. Let $\Delta_{s,t} \in \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ denote the net change in the fundamentals of stock $s$’s neighbors as of time $(t - 1)$,

$$\Delta_{s,t} = \sum_{s' \in N_s} \text{Sign}[\Delta_{s',t-1}].$$

For example, suppose that stock $s$ has $N_s = 5$ neighbors. If 2 neighbors have realized positive shocks to fundamentals by time $(t - 1)$, 1 neighbor has realized a negative shock to fundamentals by time $(t - 1)$, and the remaining 2 neighbors haven’t realized an shocks by time $(t - 1)$, then $\Delta_{s,t} = 2 \cdot (+1) + 1 \cdot (-1) + 2 \cdot 0 = +1$.

Imagine that we select a stock $s$ at random and exogenously change its fundamentals at time 1, $\Delta_{s,1} = +1$. If $s$ happens to have a neighbor, $s' \in N_s$, then the rebalancing caused by our initial shock will change the fundamentals of a second stock, $\Delta_{s',t+1} = -1$. And, if $s'$ happens to have an additional neighbor of its own, $s'' \in N_{s'} \setminus \{s\}$, then the second wave of rebalancing caused by our initial shock to stock $s$ will change the fundamentals of a third stock as well, $\Delta_{s'',t+2} = +1$. If stock $s''$ doesn’t have any additional neighbors besides stock $s'$, $N_{s''} \setminus \{s'\} = \emptyset$, then we will have triggered a rebalancing cascade of length 3 with our initial shock to a single stock’s fundamentals. Let $C_s$ denote the length of the rebalancing cascade triggered by an initial exogenous shock to stock $s$’s fundamentals,

$$C_s \overset{\text{def}}{=} |\Delta_{s,1}| + \left\{ \sum_{s' \in N_s} \left( |\Delta_{s',t+1}| + \left\{ \sum_{s'' \in N_{s'} \setminus \{s\}} (|\Delta_{s'',t+2}| + \cdots) \right\} \right) \right\}.$$

Put another way, $C_s$ is the total number of stocks whose fundamentals would change (for better or for worse) if we shocked stock $s$’s fundamentals at time $t = 1$. 7
Model Goals. Imagine you’re a market maker for stock $s_Z$. We want to answer two questions for you with this model. First, we want to characterize the expected cascade length, $\mathbb{E}[C_s]$, for a randomly selected initial stock $s$ as a function of the market’s average connectivity, $\lambda$. In other words, we want to be able to tell you if shocks on the other side of the market are likely to affect the demand for stock $s_Z$. Second, if they are, we want to be able to tell you whether it’s possible to predict how these long rebalancing cascades will affect the demand for stock $s_Z$. A long rebalancing cascade could result in either buy or sell orders for stock $s_Z$. We want to characterize how hard it is to figure out which one.

2.2 Existence of Cascades

Let’s start with the first question: how long are rebalancing cascades on average?

Low Connectivity. If stocks have less than 1 neighbor on average, $\mathbb{E}[N_s] = \lambda < 1$, then rebalancing cascades will be short. Cascades will quickly run out of nearby neighbors and peter out. Proposition 2.2a below characterizes the expected cascade length in a large market, $S \to \infty$, when the overall level of connectivity is low.

Proposition 2.2a (Cascade Length, Low Connectivity). If $\mathbb{E}[N_s] = \lambda < 1$ and $S \to \infty$, then rebalancing cascades have finite length when stocks typically have less than 1 neighbor, $\mathbb{E}[N_s] = \lambda < 1$. Right Panel. Solid line: expected fraction of the market involved in a cascade, $\mathbb{E}[C_s/S]$, as a function of market connectivity, $\lambda$, when $S \to \infty$. Dashed line: average fraction of the market involved in a cascade, $\hat{\mathbb{E}}[C_s/S]$, as a function of market connectivity, $\lambda$ in 100 simulations of a market with $S = 10^4$ stocks. Shaded region is [5%, 95%] range. Reads: “If stocks typically have more than 1 neighbor, $\mathbb{E}[N_s] = \lambda > 1$, then rebalancing cascades affect a finite fraction of an infinitely large market.”
then the expected cascade length is finite and given by:

\[ E[C_s] = \frac{1}{1-\lambda}. \]  

(4)

The left panel of Figure 2 plots this expected cascade length, \( E[C_s] \), as a function of overall market connectivity, \( \lambda \), in an infinitely large market. Increasing the number of neighbors per stock increases the expected cascade length. If stocks tend to have more neighbors, then the initial shock to stock \( s \)'s fundamentals will tend to cause more funds to rebalance. And, this second wave of rebalancing will tend to cause even more rebalancing in stock \( s \)'s neighbors. But, if stocks tend to have less than 1 neighbor, then at some point the cascade will run out of neighboring stocks—that is, cascades will have finite length. This is the region in Figure 2 to the left of \( \lambda = 1 \).

**High Connectivity.** But, if stocks have more than 1 neighbor on average, \( E[N_s] = \lambda > 1 \), then rebalancing cascades will be long. In fact, they will be infinitely long on average. And, an infinitely long rebalancing cascade, \( C_s \to \infty \), takes up a finite fraction of an infinitely large market, \( S \to \infty \). Let’s use \( \theta \in [0,1] \) to refer to the fraction of the market that’s typically affected by rebalancing cascades:

\[ \theta \overset{\text{def}}{=} \lim_{S \to \infty} \frac{E[C_s]}{S}. \]  

(5)

If \( \lambda < 1 \), then rebalancing cascades have finite length on average and \( \theta = 0 \); if \( \lambda = S \), then rebalancing cascades always affect every stock in the market and \( \theta = 1 \). Proposition 2.2b below characterizes the fraction of the market affected by rebalancing cascades in a large market, \( S \to \infty \), when the overall level of connectivity is high.

**Proposition 2.2b (Cascade Length, High Connectivity).** If \( E[N_s] = \lambda > 1 \) and \( S \to \infty \), then rebalancing cascades are infinitely long on average, taking up a fraction \( \theta \) of the market which satisfies the following equation:

\[ \theta = 1 - e^{-\lambda \theta}. \]  

(6)

The right panel of Figure 2 plots the expected fraction of the market involved in a rebalancing cascade, \( \theta = E[C_s/S|S \to \infty] \), as a function of market connectivity, \( \lambda \). When stocks have less than 1 neighbor on average, \( E[N_s] = \lambda < 1 \), rebalancing cascades have finite length and \( \theta = 0 \). When stocks have more than 1 neighbor on average, \( E[N_s] = \lambda > 1 \), rebalancing cascades are infinitely long and suddenly take up a finite fraction of the market, \( \theta > 0 \).

**Numerical Analysis.** Taken together, Propositions 2.2a and 2.2b show how the length of rebalancing cascades changes character as the average number of neighbors per stock exceeds \( E[N_s] = \lambda = 1 \). We plot the theoretically predicted cascade length (left panel) and affected
fraction (right panel) in an infinitely large market, \( S \to \infty \), with solid black lines in the left and right panels of Figure 2. And, these theoretical predictions are broadly matched in numerical simulations of a market with \( S = 10^4 \) stocks where we compute the average cascade length, \( \hat{E}[C_s] \), and the fraction of the market affected by the cascade, \( \hat{E}[C_s / S] \). We can’t get infinitely long rebalancing cascades with finitely many stocks, but we do see a qualitative jump in the length of rebalancing cascades around the threshold \( E[N_s] = \lambda = 1 \).

2.3 Cascades Generate Noise

Stock \( s_A \)'s price jumps, which causes a fund to buy stock \( s_A \) and sell stock \( s_B \), which causes a second fund to sell stock \( s_B \) and buy stock \( s_C \), which causes... We’ve just seen that these sorts of long rebalancing cascades are more likely when stocks have more neighbors. But, notice that the same rebalancing cascade affects different stocks in different ways. For instance, \( s_B \) realized a sell order while \( s_C \) realized a buy order. We now show that, in a large market, it’s computationally infeasible to predict how a long rebalancing cascade will affect the demand for an a randomly selected stock.

Will \( s_Z \) Be Affected? The top panel of Figure 3 depicts 9 realizations of rebalancing cascades on a random network with \( S = 100 \) stocks where stocks have \( E[N_s] = \lambda = 1.1 \) neighbors on average. All 9 rebalancing cascades start with an exogenous initial shock to the same stock’s fundamentals. This initial stock, stock \( s_A \), is denoted by the big circle in the upper left of each panel. The rebalancing cascade that this shock to stock \( s_A \) sets off is denoted by the thick edges. We’re trying to predict how these rebalancing cascades will affect the demand for stock \( s_Z \) (large square, lower right). Sometimes the rebalancing cascades die out quickly and don’t affect the demand for stock \( s_Z \) at all. For example, the rebalancing cascade only involves \( C_A = 2 \) stocks in Panel (7) and \( C_A = 3 \) stocks in Panel (2). But, in Panels (5), (6), and (9) there are long rebalancing cascades that eventually affect the demand for stock \( s_Z \).

If all we had to do was predict whether or not stock \( s_Z \) will be affected by a long rebalancing cascade starting with stock \( s_A \), then we would be facing a relatively easy task. It’s easy to look at the top panel of Figure 3 and see whether the cascade starting with stock \( s_A \) eventually affects stock \( s_Z \). Lemma 2.3 below characterizes the typical length of the shortest path between two randomly selected stocks in a rebalancing cascade.

**Lemma 2.3** (Shortest Path). When \( \lambda > 1 \), the average length of the shortest path between two stocks involved in a rebalancing cascade, \( \ell \), is given by:

\[
\ell = \frac{\log(S) + \log(\theta)}{\log(\lambda)}.
\]
(a) Rebalancing Cascades Starting with Stock $s_A$

(1) (2) (3)
(4) (5) (6)
(7) (8) (9)

(b) Effect of Cascades on the Demand for Each Stock

(1) (2) (3)
(4) (5) (6)
(7) (8) (9)

Figure 3: Top Panel. 9 realizations of rebalancing cascades on a random network with $S = 100$ stocks where stocks have $\lambda = 1.1$ neighbors on average. Rebalancing cascades always start with the same stock $s_A$ (large circle, upper left). We’re trying to predict the effect of these rebalancing cascades on the demand for stock $s_Z$ (large square, lower right). Bottom Panel. The same 9 rebalancing cascades color coded to indicate whether each stock in the rebalancing cascade was hit with a buy order (blue) or a sell order (red).
Checking whether stock \( s_Z \) will be affected by a rebalancing cascade starting with stock \( s_A \) means starting down the list of all paths of length \( O(\ell) \) that terminate in stock \( s_Z \) and checking if at least one contains stock \( s_A \).

**How Will \( s_Z \) Be Affected?** Unfortunately, this isn’t all we have to do. The bottom panel of Figure 3 plots the exact same rebalancing cascades color coded to indicate whether each stock in the rebalancing cascade was hit with a buy order (blue) or a sell order (red). Notice that the long rebalancing cascades in Panels (5) and (9) result in sell orders (red) for stock \( s_Z \); whereas, the long rebalancing cascade in Panel (6) results in a buy order (blue) for stock \( s_Z \). While it’s easy to look at the top of Figure 3 and check whether the cascade starting with stock \( s_A \) eventually affects stock \( s_Z \), it’s hard to tell whether the effect will be positive or negative. There’s no obvious feature of Panel (5) in Figure 3a that suggests a sell order for stock \( s_Z \). There’s no smoking gun in Panel (6) in Figure 3a that suggests a buy order for stock \( s_Z \). Each rebalancing cascade is perfectly deterministic, but if you want to figure out how the rebalancing cascade will affect the demand for stock \( s_Z \), then you’re going to have to see how the rebalancing cascade plays out. The total number of people in France at any given instant isn’t random, but whether or not this number is even or odd may as well be.

**Computational Complexity.** Panel (6) of Figure 3 gives an example of why it’s so hard to predict the effect of a long rebalancing cascade on the demand for stock \( s_Z \). The left-most column of Figure 4 replicates Panel (6) from Figure 3. The right 3 columns then show 3 different paths that a rebalancing cascade could take from stock \( s_A \) to stock \( s_Z \). Paths 1 and 3 start out going in opposite directions from \( s_A \). Path 2 starts out going in the same direction as path 1, but it jumps over to following the same route as path 3 at around the half-way point. Let’s refer to the stock where paths 1 and 2 diverge “stock 1” and the stock where paths 2 and 3 recombine as “stock 6”.

Notice that something interesting happens when paths 2 and 3 reach stock 6. Path 2 includes a fund that rebalances and submits a sell order for stock 6 (red); whereas, path 3 includes a fund that rebalances and submits a buy order for stock 6 (blue). Thus, these 2 paths cancel each other out when they recombine. This sort of cancelling happens whenever there are 3 different paths that combine to create an odd-length cycle. For instance, if we count the number of stocks in the loop from stock 1 to stock 6 and back, we see that there are 11 stocks. Thus, to compute the net effect of a long rebalancing cascade starting with stock \( s_A \) on the demand for stock \( s_Z \) you have to do much more than find a single path from stock \( s_A \) to stock \( s_Z \). You have to find all such paths and figure out which ones cancel each other out by finding every single odd-length cycle in the graph. Predicting the net effect of a long rebalancing cascade on any particular stock’s demand is difficult because it means keeping
track of where the many different branches of the cascade cancel each other out and where they reinforce one another. It means keeping track of the global structure of the market.

**Proposition 2.3** (Computational Complexity). *The problem of predicting the effect of a long rebalancing cascade starting with stock $s_A$ on the demand for stock $s_Z$ is NP Complete.*

## 3 Empirical Analysis

We just saw that in a random-networks model long, unpredictable rebalancing cascades will occur whenever stocks are connected by many different threshold-based trading strategies. But, is there any evidence that these sorts of long rebalancing cascades actually occur in real-world markets? And, if so, do market makers treat the resulting demand as unpredictable noise? We examine data on the holdings of exchange-traded funds (ETFs) and find that the answer to both these questions is an emphatic “Yes.”
3.1 ETF Market

To test the hypothesis that long rebalancing cascades generate noise, we need data on the portfolio holdings of many funds that use a diverse group of simple, threshold-based rebalancing rules. Morningstar provides data on the end-of-day portfolio holdings of ETFs, a group of funds that track the returns of pre-specified benchmarks and have shares that trade on an exchange just like those of any other common stock. We now describe why the ETF market satisfies all of the above criteria and provides an ideal empirical laboratory.

Many Funds. If the rebalancing rules of a particular group of funds are going to generate long cascades, then this has to be a large group of funds since each link in a cascade represents a different fund’s rebalancing decision. The ETF market is certainly large. In fact, as of July 2015, “the amount of dollars exchanging hands through ETFs is now more than the U.S. gross domestic product, which stands at $17.4tr.”\textsuperscript{1} Our data from Morningstar covers the period from January 1st, 2005 to December 31st, 2011; but, ETFs don’t start reporting to Morningstar on a daily basis until mid 2009. So, in our main analysis, we restrict our sample to the subset of our data from January 1st, 2010 to December 31st, 2011 where we have daily data.

Table 1 gives summary statistics describing the number of ETFs reporting to Morningstar each month. Row 1 of Panel \((a)\) shows that during our sample period there were 136.63 ETFs reporting to Morningstar each month on average. There was a lot of growth in the ETF market during this time period. Figure 5 plots the number of ETFs reporting to Morningstar, which grew from 44 ETFs in January 2005 to 220 ETFs in December 2011. Row 1 of Panel \((b)\) in Table 1 shows that the typical ETF in our sample holds somewhere between \(10^{2.90} = 100\) stocks and \(10^{2.78} = 600\) stocks. And, Row 2 of Panel \((b)\) reveals that the market value of an ETF’s portfolio holdings usually falls somewhere between \(\$1 \times 10^{7.47} = \$30m\) and \(\$1 \times 10^{9.12} = \$1.3b\).

Simple Rules. In Section 2 above, we showed theoretically that long rebalancing cascades can generate unpredictable demand shocks. In the model, these demand shocks are unpredictable because long rebalancing cascades are complicated objects not because the individual rebalancing decisions were complicated. So, for our empirical analysis, we want data on a group of funds following simple, pre-specified rebalancing rules. ETFs are just such a group. In fact, there are really only 2 reasons why ETFs trade: creations/redemptions and benchmark-related rebalancing.

Let’s look at creations and redemptions first. The company running an ETF (its “sponsor”) has an obligation to create or redeem shares of the ETF at the end-of-day market value

\textsuperscript{1} Balchunas, E. *ETF Trading Has Eclipsed US GDP*. Bloomberg, 07/30/15. https://goo.gl/5EA6NH
of its stated benchmark. If an ETF’s price is higher than the end-of-day market value of its benchmark, then an arbitrageur can sell shares of the ETF back to its sponsor and use the proceeds to buy shares of the assets in the benchmark. Conversely, if an ETF’s price is lower than the end-of-day market value of its benchmark, then an arbitrageur can sell shares of the assets in the benchmark and use the proceeds to buy shares of the ETF from its sponsor. Thus, ETF managers will always hold a basket of securities that closely mirrors the end-of-day market value of their stated benchmark. An arbitrageur will profit and the sponsor will lose money whenever there’s a difference between the market value of an ETF’s end-of-day holdings and the market value of that ETF’s stated benchmark.\footnote{See \url{https://goo.gl/TuhjAO} for a fun interactive illustration.}

Following this logic through to its natural conclusion, if arbitrageurs are constantly asking an ETF sponsor to create or redeem lots of shares, then the sponsor must be losing lots of money. So, just like you’d expect, creations and redemptions are only a small fraction of daily trading volume for ETFs, and these trades involve less than 0.5\% percent of ETFs’ net assets (Investment Company Institute, 2015). Instead, ETF trading volume primarily comes from rebalancing activity just prior to market close. This end-of-day trading is how ETF sponsors make sure that there is very little difference between the market value their end-of-day holdings and the market value of their stated benchmark. In our empirical analysis below, we’re going to make use of the fact that most of ETFs’ trading volume is due to rebalancing activity in the final 20 minutes before market close.

In the data from Morningstar, we observe both assets under management and portfolio weights for each ETF on a daily basis. Let $aum_{f,t}$ denote an ETF’s assets under management on day $t$, and let $\omega_{f,s,t}$ be the fraction of this money that the ETF invests in stock $s$ on day

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{chart.png}
\caption{Red circles: total number of ETFs that report their portfolio holdings to Morningstar each month. Blue triangles: number of unique benchmarks tracked by ETFs reporting to Morningstar each month. Reads: “The number of ETFs reporting to Morningstar grew from 44 ETFs in January 2005 to 220 ETFs in December 2011. What’s more, these new ETFs aren’t very similar to one another: the 220 ETFs in December 2011 track 191 different benchmarks.”}
\end{figure}
Thus, if \( p_{s,t} \) is the price of stock \( s \) on day \( t \), then the actual number of shares of stock \( s \) that the \( f \)th ETF holds on day \( t \), \( x_{f,s,t} \), is given by:

\[
x_{f,s,t} = \left( \omega_{f,s,t} \cdot \text{aum}_{f,t} \right) / p_{s,t}.
\]

And, the total ETF-trading volume for stock \( s \) on day \( t \) is given by \( \sum_{f=1}^{F} |x_{f,s,t} - x_{f,s,t-1}| \). This calculation includes both rebalancing and trading due to inflows and outflows.

But, in this paper, we are interested in ETF-rebalancing volume. To compute this quantity, we use a slight variant on the above calculation. Notice that if we were to use the \( f \)th ETF’s portfolio weights on the previous day in Equation (8), then we’d have the ETF’s predicted holdings of stock \( s \) on day \( t \) given the ETF’s realized inflows and outflows:

\[
\bar{x}_{f,s,t} = \left( \omega_{f,s,t-1} \cdot \text{aum}_{f,t} \right) / p_{s,t}.
\]

If money pours into ETF \( f \) on day \( t \) (\( \text{aum}_{f,t} \gg \text{aum}_{f,t-1} \)), then it’s going to have to buy shares of stock \( s \), but this trading volume won’t be due to rebalancing. So, to estimate the amount of ETF-rebalancing volume for stock \( s \) on day \( t \), we compute

\[
\text{etfRebalancing}_{s,t} \overset{\text{def}}{=} \sum_{f=1}^{F} |x_{f,s,t} - \bar{x}_{f,s,t}|
\]

rather than \( \sum_{f=1}^{F} |x_{f,s,t} - x_{f,s,t-1}| \).

**Diverse Group.** If all the funds in our data were using the same rebalancing rules, then they would all rebalance in the exact same way in response to an initial price shock. In the language of the theoretical analysis from Section 2, this would mean that each stock could only be connected to at most \( E[N_s] = \lambda \leq 1 \) neighbors via a portfolio rebalancing rule. And, if this were the case, then Proposition 2.2a says that long rebalancing cascades would be impossible. Thus, to have long ETF-rebalancing cascades it is important that the funds we look at follow a diverse group of rebalancing rules. ETFs certainly fit this description.

When people think of an ETF, they usually think of a passively managed ETF, like the SPY, that tracks a broad value-weighted market index, like the S&P 500 index. After all, the SPY is the world’s most-traded security, and prior to January 2008 all ETFs did in fact look a lot like the SPY. They all had to mirror the returns of a pre-existing benchmark index, like the S&P 500. But, in early 2008, the SEC changed its guidelines so that ETFs could track a self-defined benchmark. After this change, Invesco PowerShares was free to create an ETF that tracked the returns of the quintile of S&P 500 stocks with the lowest historical volatility even though there was no pre-existing low-volatility S&P 500 index. All Invesco had to do was promise to announce the identities and weights of the constituents in its self-defined low-volatility benchmark one day in advance.

Because of this 2008 rule change, ETFs now follow a diverse group of self-defined bench-
**Figure 6: Top Panels.** Each point represents a stock. Large circle: stock $s_A$, which is the target in an M&A announcement at time $\tau_A$. Large square: stock $s_Z$, which is a stock that’s unrelated to stock $s_A$. In column (1), stock $s_Z$ is connected to $N_Z = 1$ neighboring stock via a portfolio rebalancing rule, denoted by a black line emanating from stock $s_Z$. In column (2), stock $s_Z$ is connected to $N_Z = 2$ neighboring stocks via portfolio rebalancing rules. In column (3) stock $N_Z = 3$, and in column (4) stock $N_Z = 4$. When stocks are connected to $E[N_s] = \lambda = 1.10$ neighbors on average, Median($N_s$) = 1. So, if $\lambda = 1.10$, then stock $s_Z$ is held by an above-median number of ETFs in columns (2), (3), and (4). Bottom Panels. Sample averages from $10^4$ simulations of a market with $S = 100$ stocks where each stock has $E[N_s] = \lambda = 1.10$ neighbors on average and stock $s_Z$ has 1, 2, 3, or 4 neighbors. $Pr(Hit)$: fraction of the $10^4$ simulations where stock $s_Z$ was hit by the rebalancing cascade starting with stock $s_A$. $Pr(Buy|Hit)$: in the set of simulations where stock $s_Z$ was hit by the long rebalancing cascade, this is the fraction of the time that the cascade resulted in a net buy order for stock $s_Z$.

Figure 5 shows that the 220 ETFs reporting to Morningstar in our data from December 2011 were tracking 191 different benchmarks, ranging from the completely boring (S&P 500 index; technology-sector index) to the somewhat interesting (S&P 500 Low-Volatility ETF; Russell 1000 Low-Beta ETF) to the downright niche (Drone Economy Strategy ETF; 3D Printing ETF; Baptist Values ETF). To be sure, niche funds tend to be smaller. But, even the rebalancing activity of small ETFs can affect a stock’s fundamentals because ETFs do all of their rebalancing during the final 20 minutes of the trading day. “The more indexing there is, generally, the more close-trading there will be,” said Paul Whitehead, who helps oversee trading for index funds at the $4.8tr fund manager, BlackRock Inc.”

**Thresholds.** Finally, let’s think about the role that threshold-based rebalancing rules play in our analysis. It would still be possible to have long rebalancing cascades without

---

3Strumpf, D. *Stock-Market Traders Pile In at the Close.* WSJ, 05/27/15. [https://goo.gl/ErQZA8](https://goo.gl/ErQZA8)
thresholds. For example, even if portfolio rebalancing’s effect on market fundamentals is continuous, a 1-time initial shock can have long-range effects in a simple AR(1) model, $\Delta_t = -\lambda \cdot \Delta_{t-1} + \epsilon_t$, so long as $\lambda \approx 1$. But, notice that the effect of an initial shock is half as strong after $\log \lambda(\frac{1}{2})$ trades in this AR(1) model. This is an important observation. Without thresholds, the long-range effects of a rebalancing cascade are easy to predict because longer cascades are weaker cascades. Without thresholds, market makers would only have to worry about the most direct routes that a rebalancing cascade could take from an initial shock to stock $s_Z$ because longer routes would have smaller effects on the demand for stock $s_Z$. Thus, threshold-based rebalancing rules are important because they break this link between cascade length and cascade strength. Each domino falls with the same force regardless of whether it’s the first, the last, or anywhere in between.

Row 5 of Panel (a) in Table 1 shows that 68% of all ETFs reporting Morningstar track benchmarks that involve at least 2 different thresholds. For example, the SPY tracks a benchmark (the largest 500 US stocks) with a single threshold. An arbitrarily small increase in the size of the 501st largest stock can push it into this list. By contrast, the PowerShares S&P 500 Low-Volatility ETF tracks a benchmark (the quintile of S&P 500 stocks with the lowest volatility) with 2 thresholds. Just like before, an arbitrarily small change in a stock’s market cap can push it from being the 501st largest company to being the 500th largest company. But now, an arbitrarily small difference in volatilities can also move a stock from having the 101st to having the 100th lowest volatility among S&P 500 stocks.

### 3.2 Empirical Strategy

To test the hypothesis that long rebalancing cascades generate noise in financial markets, we need to verify 2 things: that long rebalancing cascades exist and that market makers treat the demand coming from these cascades as unpredictable noise. We now outline our strategy for verifying these 2 predictions.

**Initial Shocks.** Long rebalancing cascades need to start with some sort of initial price shock. The longest journey begins with a single step. Something has to knock over the first domino. We look at initial price shocks that come from M&A announcements, and we refer to the stock that’s the target of the initial M&A announcement as stock $s_A$. Our source for M&A deals is Thomson Financial. We use all deals with an announcement date between January 1st, 2010 and December 31st, 2011 where the target is a public company. Row 1 of Panel (a) in Table 2 describes the number of M&A announcements we observe each quarter. We represent stock $s_A$, the stock which realizes the initial M&A-related price shock, with a large circle in Figure 6. M&A deals are a natural choice because, in the words of Andrade
et al. (2001), “a profusion of event studies has demonstrated that mergers seem to create shareholder value, with most of the gains accruing to the target company.” While acquirers do not choose their M&A targets at random, the exact day that a deal is announced can be taken as random. Let \( \tau_A \) denote the day of the M&A announcement in which stock \( s_A \) is the target firm. We use the indicator variable, \( \text{afterAnnouncement}_{A,t} \), to flag the 10 day period following the M&A-related price shock to stock \( s_A \):

\[
\text{afterAnnouncement}_{A,t} \overset{\text{def}}{=} \begin{cases} 
1 & \text{if } t \in [\tau_A, \tau_A + 10), \\
0 & \text{otherwise}. 
\end{cases}
\] (11)

**Unrelated Stocks.** We are interested in whether or not these initial M&A-related price shocks set off long, unpredictable ETF-rebalancing cascades that affect the demand for unrelated stocks on the other side of the market. For each M&A deal there will be many unrelated stocks. We refer to an arbitrary one of these unrelated stocks as stock \( s_Z \). Stock \( s_Z \) can be any stock that isn’t in the same industry as stock \( s_A \) and isn’t held by any of the same ETFs that also hold stock \( s_A \). Row 2 of Panel (a) in Table 2 describes the average number of unrelated stocks on the other side of the market to M&A announcements each quarter. We represent stock \( s_Z \) with a large square in Figure 6.

From the theoretical analysis in Section 2, we know that a stock is more likely to be hit by a long rebalancing cascade when it is involved in more rebalancing rules. To illustrate, the bars labeled \( \text{Pr(Hit)} \) in Figure 6 report the fraction of the time that stock \( s_Z \) is hit by a long rebalancing cascade starting with stock \( s_A \) in \( 10^4 \) simulations of a market with \( S = 100 \) stocks where each stock has \( E[N_s] = \lambda = 1.10 \) neighbors on average. As stock \( s_Z \) moves from having only \( N_Z = 1 \) neighbor to having \( N_Z = 4 \) neighbors, the probability that it’s hit by a rebalancing cascade that originates with stock \( s_A \) nearly doubles, moving from 11.0% to 21.0%. At any point in time, we refer to the unrelated stocks that are held by an above-median number of ETFs for the quarter as our treatment group:

\[
isTreated_{Z,t} \overset{\text{def}}{=} \begin{cases} 
1 & \text{if } s_Z \text{ is held by above-median number of ETFs in quarter, and} \\
0 & \text{otherwise.} 
\end{cases}
\] (12)

In Figure 6, stock \( isTreated_{Z,t} = 1 \) in columns (2), (3), and (4).

**Existence of Cascades.** We verify the existence of long ETF-rebalancing cascades by showing that ETF trading activity for stock \( s_Z \) increases in the days after an M&A announcement about an unrelated stock \( s_A \) at times when stock \( s_Z \) happens to be held by more ETFs. This relationship between ETF-rebalancing volume on one side of the market and shocks to unrelated stocks on the other side of the market can’t exist in a world without
long rebalancing cascades. To estimate this effect, we collect 60 days of data around each M&A announcement, \([\tau_A - 50, \tau_A + 10]\), for each unrelated stock on the other side of the market. Then, we run the following regression where the \(A \rightarrow Z\) indexing captures the fact that a single stock \(s_Z\) might be on the other side of the market from 2 M&A announcements:

\[
\log(\text{etfRebalancing}_{A \rightarrow Z, t}) = \hat{\alpha}_A + \hat{\alpha}_Z + \hat{\alpha}_t \\
+ \hat{\beta} \times \text{afterAnnouncement}_{A, t} \\
+ \hat{\gamma} \times \{\text{afterAnnouncement}_{A, t} \times \text{isTreated}_{Z, t}\} \\
+ \epsilon_{A \rightarrow Z, t}.
\]

(13)

\(\log(\text{etfRebalancing}_{A \rightarrow Z, t})\) is the log of the ETF-rebalancing volume for stock \(s_Z\) on day \(t\). \(\hat{\alpha}_A, \hat{\alpha}_Z, \) and \(\hat{\alpha}_t\) are event, stock, and year-month fixed effects. The coefficient \(\hat{\beta}\) captures how much ETF-rebalancing volume increases in the days after an M&A announcement about an unrelated stock. The coefficient \(\hat{\gamma}\) is really what we’re interested in. It captures how much more ETF-rebalancing volume increases when the same stock \(s_Z\) is held by many ETFs. We should only estimate a \(\hat{\gamma} > 0\) if there are long ETF-rebalancing cascades. Note that, because we are using stock fixed effects, our results can’t be explained by differences in ETF trading across stocks. We’re looking at the percentage increase in ETF-rebalancing volume for the same stock at times when that stock just happens to be held by a few more ETFs precisely as illustrated in Figure 6.

**Cascades Generate Noise.** In Kyle-type models, market makers learn about informed traders’ private information by observing aggregate demand, which is the sum of informed-trader demand and noise. The standard result from this literature is that, if market makers face more demand noise, then they will have a harder time learning informed traders’ private information, so informed traders will trade more aggressively. So, we show that market makers treat the demand from long ETF-rebalancing cascades as noise by showing that informed traders are more likely to trade stock \(s_Z\) in the days after an M&A announcement about an unrelated stock \(s_A\) at times when stock \(s_Z\) happens to be held by more ETFs. This relationship between informed demand on one side of the market and shocks to unrelated stocks on the other side of the market shouldn’t exist in a world without long ETF-rebalancing cascades that market makers treat as noise.

We follow Collin-Dufresne and Fos (2015) and use Rule 13(d) of the 1934 Securities Exchange Act to identify demand from informed traders. This rule says that shareholders have to let the Securities and Exchange Commission (SEC) know whenever they acquire 5% or more of a publicly traded company that they have a voting interest in. We use the indicator variable, \(\text{hasInformedTrading}_{A \rightarrow Z, t}\), to flag days when unrelated stocks are traded
by Schedule 13D filers in the days after an M&A-related price shock to stock $s_A$:

$$\text{hasInformedTrading}_{A\rightarrow Z,t} \overset{\text{def}}{=} \begin{cases} 1 & \text{if informed traders trade stock } s_Z \text{ on day } t, \\ 0 & \text{otherwise}. \end{cases}$$ (14)

We then run the exact same sort of regression as in Equation (13) to test whether informed traders are more likely to trade stock $s_Z$ in the days after an M&A announcement about an unrelated stock $s_A$ at times when stock $s_Z$ happens to be held by more ETFs:

$$\text{hasInformedTrading}_{A\rightarrow Z,t} = \hat{\alpha}_A + \hat{\alpha}_Z + \hat{\alpha}_t + \hat{\beta} \times \text{afterAnnouncement}_{A,t} + \hat{\gamma} \times \{\text{afterAnnouncement}_{A,t} \times \text{isTreated}_{Z,t}\} + \epsilon_{A\rightarrow Z,t}. \quad (15)$$

Instead of looking at percentage changes in ETF-rebalancing volume for stock $s_Z$, we now look at whether or not informed traders are more likely to trade stock $s_Z$. So, the coefficient $\hat{\gamma}$ is the coefficient of interest just like before. $\hat{\gamma}$ captures how much more likely informed traders are to trade stock $s_Z$ in the aftermath of an M&A-related price shock to stock $s_A$ when stock $s_Z$ happens to be held by many ETFs. We should only estimate a $\hat{\gamma} > 0$ if market makers treat the demand coming from long ETF-rebalancing cascades as noise.

**The Mechanism.** Finally, we make use of a key institutional detail about how ETFs trade to show that ETF-rebalancing activity is really what’s driving our results. Specifically, in Section 3.1 we described how most ETF-trading volume comes from rebalancing at the end of each trading day. So, as a proxy for whether the M&A-related price shock to stock $s_A$ had any impact on ETF-rebalancing activity, we check whether stock $s_A$ saw a spike in end-of-day trading volume since this spike in volume is likely due to ETF-rebalancing activity. Then, we check whether our regression results from Equations (13) and (15) are stronger for M&A-related price shocks where stock $s_A$ had a spike in end-of-day trading volume. We say that stock $s_A$ has lots of end-of-day trading volume on its announcement day, $\tau_A$, if it has an above-median fraction of its trading volume occurring within 20 minutes of the market close on its announcement day, $\tau_A$:

$$\text{endOfDayActivity}_{s,t} \overset{\text{def}}{=} \begin{cases} \text{Hi} & \text{if } \frac{\text{volumeInLast20Min}_{s,t}}{\text{volume}_{s,t}} \text{ is above median for quarter, and} \\ \text{Lo} & \text{otherwise}. \end{cases} \quad (16)$$

### 3.3 Main Results

We find that, when stock $s_A$ realizes a price shock, stock $s_Z$ on the other side of the market only sees increased ETF-rebalancing volume at times when it’s held by many different ETFs.
and thus more susceptible to long ETF-rebalancing cascades. And, we document that stock $s_Z$’s informed traders trade more aggressively at precisely these same moments, suggesting that stock $s_Z$’s market makers view the erratic, non-fundamental demand coming from long ETF-rebalancing cascades as noise.

**Existence of Cascades.** If there aren’t long ETF-rebalancing cascades, then there’s no reason that price shocks on one side of the market should lead to more ETF-rebalancing volume for unrelated stocks on the other side of the market that are held by lots of ETFs. This is the simple idea behind our empirical test for the existence of long ETF-rebalancing cascades in Equation (13). The left-most column of Table 3 displays the estimated $\hat{\gamma}$ from Equation (13), which provides strong evidence in favor of the existence of long ETF-rebalancing cascades. In the 10 days after an M&A-related price to stock $s_A$, we estimate that ETF-rebalancing volume for an unrelated stock $s_Z$ will grow by 14% more when stock $s_Z$ is held by many ETFs. This point estimate is statistically significant even when clustering the standard errors by stock. And, because we use $s_Z$-specific fixed effects, our results can’t be explained by differences in ETF trading across stocks. We’re looking at the percentage increase in ETF rebalancing for the same stock at times when that stock just happens to be held by a few more ETFs.

**Cascades Generate Noise.** If market makers don’t treat the demand coming from long ETF-rebalancing cascades as noise, then there’s no reason that price shocks on one side of the market should lead to additional informed trading in unrelated stocks on the other side of the market that are held by lots of ETFs. This is the simple idea behind our empirical test that market makers treat demand from long ETF-rebalancing cascades as noise in Equation (15). The left-most column of Table 4 displays the estimated $\hat{\gamma}$ from Equation (15), which provides strong evidence that market makers treat the demand coming from these long ETF-rebalancing cascades as noise. In the 10 days after an M&A-related price to stock $s_A$, we estimate that stock $s_Z$ Schedule 13D filers are 0.04% more likely to trade stock $s_Z$ when it is held by many ETFs. This point estimate is statistically significant even when clustering the standard errors by stock. And, although the point estimate seems small at first, it is actually quite economically significant when you realize that Schedule 13D trading is quite uncommon. Table 2 reports that Schedule 13D filers trade on 0.25% of all trading days in our sample. So, our estimated 0.04% increase represents a 16% increase in the rate of informed trading, which is a big effect.

**The Right Mechanism.** Finally, the right-most columns in Tables 3 and 4 show how these two results change when we split our sample of initial M&A-related price shocks. Most
ETF trading volume comes from rebalancing at the end of each trading day. So, if stock \( s_A \) realized a spike in end-of-day trading volume on the day of its M&A announcement, \( \tau_A \), then it’s more likely that the M&A-related price shock forced some ETFs to rebalance. The middle column in Table 3 (labeled “Hi”) shows the estimated coefficient \( \hat{\gamma} \) from Equation (13) when we restrict our sample to M&A targets that realized a spike in end-of-day trading volume on the day of their announcement. In the 10 days after an M&A-related price shock to stock \( s_A \) that leads to a spike in end-of-day trading volume, we estimate that ETF trading volume for an unrelated stock \( s_Z \) will grow by 17\% more when stock \( s_Z \) is held by many ETFs. The right-most column in Table 3 (labeled “Lo”) shows the estimated \( \hat{\gamma} \) from Equation (13) when we restrict our sample to M&A targets that didn’t realize a spike in end-of-day trading volume on the announcement day. If the M&A-related price to stock \( s_A \) doesn’t generate a spike in end-of-day trading volume, then we estimate that ETF-rebalancing volume for an unrelated stock \( s_Z \) will only grow by 13\% more when stock \( s_Z \) is held by many ETFs. That is, our estimate of \( \hat{\gamma} \) shrinks when we look at the initial price shocks that least likely to have caused an ETF-rebalancing cascade. This difference between the estimated \( \hat{\gamma} \) in columns “Hi” and “Lo” is statistically significant at 1\% level.

We find similar effects when we examine the evidence that these long ETF-rebalancing cascades generate noise. The middle column in Table 4 shows that in the 10 days after an M&A-related price shock to stock \( s_A \), the chances of informed trading for an unrelated stock \( s_Z \) will grow by 0.06\%pt more when stock \( s_Z \) is held by many ETFs. By contrast, the right-most column in Table 3 shows that the chances of informed trading for an unrelated stock \( s_Z \) will not change significantly when stock \( s_Z \) is held by few ETFs. Again, this difference between the estimated \( \hat{\gamma} \) in columns “Hi” and “Lo” is statistically significant at 1\% level. That is, our estimate of \( \hat{\gamma} \) shrinks when we look at the initial price shocks that least likely to have caused an ETF-rebalancing cascade. Overall, long ETF-rebalancing cascades appear to be the mechanism driving our results.

### 3.4 Robustness Checks

We examine a pair of additional tests to reinforce our main results.

*No Evidence of Selection.* In our main analysis, we say that stock \( s_Z \) is treated if it is held by an above-median number of ETFs for the quarter. We pick this treatment because stock \( s_Z \) is more likely to be hit by a long ETF-rebalancing cascade when it is held by many ETFs. And, we find that treated stocks realize larger jumps in their ETF-rebalancing volume and more informed trading in the wake of M&A-related price shocks on the other side of the market. A natural question when first reading the informed-trading results is, “Maybe the
results have nothing to do with ETF trading? Perhaps informed traders just trade stock \( s_Z \) differently when it’s held by lots of ETFs regardless of whether or not stock \( s_Z \) realizes a long ETF-rebalancing cascade?” It could be that we’re selecting times when Schedule 13D filers are going to be behaving fundamentally different if we only look at stock \( s_Z \) when it’s held by many ETFs.

This is a good question. And, we estimate the following regression to address it:

\[
\frac{\text{informedVolume}_{s,t}}{\text{sharesOutstanding}_{s,t}} = \hat{\beta} \times \left( \frac{\text{etfVolume}_{s,t}}{\text{sharesOutstanding}_{s,t}} \right) \\
+ \hat{\gamma} \times \log(\text{etfsHoldingStock}_{s,t}) \\
+ \hat{\delta} \times \left\{ \log(\text{etfsHoldingStock}_{s,t}) \times \left( \frac{\text{etfVolume}_{s,t}}{\text{sharesOutstanding}_{s,t}} \right) \right\} \\
+ \hat{\alpha}_t + \epsilon_{s,t}.
\]  

(17)

The left-hand-side variable, \( \frac{\text{informedVolume}_{s,t}}{\text{sharesOutstanding}_{s,t}} \), denotes trading volume in stock \( s \) coming from Schedule 13d filers normalized by the total number of shares outstanding. \( \frac{\text{etfVolume}_{s,t}}{\text{sharesOutstanding}_{s,t}} \) denotes the ETF trading volume in the same stock, again normalized by the total number of shares outstanding. And, \( \log(\text{etfsHoldingStock}_{s,t}) \) is the log of the number of ETFs that held stock \( s \). If Schedule 13d filers behaved differently when stock \( s_Z \) is held by more ETFs, then we would expect to estimate \( \hat{\gamma} \neq 0 \). But, when we examine the data, we find that \( \hat{\gamma} \) is a precisely estimated 0 as reported in Table 5.

**Informed Traders Respond to Liquidity.** Here’s another question that often comes up when people first see the main results: “If market makers can’t figure out these long ETF-rebalancing cascades, then how do Schedule 13d filers do it?” The short answer is: “They don’t have to.” If market makers treat the demand coming from long ETF-rebalancing cascades as noise, then stock \( s_Z \) should become more liquid following a price shock to stock \( s_A \) on the other side of the market. All stock \( s_Z \) Schedule 13d filers are doing is taking advantage of this liquidity. Schedule 13d filers don’t need to know or care where this increase in liquidity is coming from.

To show that this liquidity channel is what’s driving our results, we replace the left-hand-side variable in our main regression specifications from Equations (13) and (15) with illiquidity measures for stock \( s_Z \). First, we look at the Amihud (2002) illiquidity measure:

\[
\text{amihudIlliquidity}_{A \rightarrow Z,t} = \hat{\alpha}_A + \hat{\alpha}_Z + \hat{\alpha}_t \\
+ \hat{\beta} \times \text{afterAnnouncement}_{A,t} \\
+ \hat{\gamma} \times \{ \text{afterAnnouncement}_{A,t} \times \text{isTreated}_{Z,t} \} \\
+ \epsilon_{A \rightarrow Z,t}.
\]  

(18)

The coefficient \( \hat{\gamma} \) will tell us how much less liquid \( s_Z \) is following a price shock on the other
side of the market when its held by many ETFs. If informed traders are responding to liquidity, then we should expect to estimate a negative value for $\hat{\gamma}$. And, in Panel (a) of Table 6, this is exactly what we find. We report similar results in Panel (b) of Table 6 where we use measure the illiquidity of stock $s_Z$ with its average bid-ask spread:

$$
\text{bidAskSpread}_{A \rightarrow Z,t} = \hat{\alpha}_A + \hat{\alpha}_Z + \hat{\alpha}_t \\
+ \hat{\beta} \times \text{afterAnnouncement}_{A,t} \\
+ \hat{\gamma} \times \{ \text{afterAnnouncement}_{A,t} \times \text{isTreated}_{Z,t} \} \\
+ \epsilon_{A \rightarrow Z,t}.
$$

\[19\]

4 Conclusion

This paper proposes that long rebalancing cascades generate noise in financial markets. We show theoretically that, when funds follow many different threshold-based trading strategies, a small change in stock $s_A$’s fundamentals can trigger a long rebalancing cascade that eventually affects the demand for unrelated stocks on the other side of the market, like stock $s_Z$. And, we prove that, in a large market, it’s computationally infeasible to predict whether long rebalancing cascades will result in buy or sell orders for stock $s_Z$. The best that a market maker can do is compute stock $s_Z$’s susceptibility to these erratic, non-fundamental shocks—its susceptibility to this demand noise.

We use data on the ETF market to empirically verify this proposal. But, ETFs aren’t the only funds that use threshold-based trading rules. For example, every hedge fund following a momentum strategy is using a threshold-based trading strategy. They’re buying the 30% of stocks with the highest past return and selling the 30% of stocks with the lowest. These sorts of threshold-based trading strategies are everywhere because “any predictive regression can be expressed as a portfolio sort (Pedersen, 2015).” The results in this paper suggest that the collective behavior of all these threshold-based trading strategies generates the noise that makes financial markets possible.
References


A Proofs

Proof (Lemma 2.1). If stock \( s \) is connected to any stock \( s' \in \mathcal{S} \) with probability \( \lambda/s \), then the probability that stock \( s \) is connected to \( N_s = n \) neighbors is governed by:

\[
\Pr[N_s = n] = \text{Bin}(n; \lambda/s) = \binom{S}{n} \cdot \left(\frac{\lambda}{s}\right)^n \cdot \left(1 - \frac{\lambda}{s}\right)^{S-n}.
\]  

(20)

And, we have that \( \lim_{S \to \infty} \text{Bin}(\lambda/s) = \text{Pois}(\lambda) \).

The proofs of Proposition 2.2a, Proposition 2.2b, and Lemma 2.3 use probability-generating functions. See Graham et al. (1989) for more details.

Proof (Proposition 2.2a). Let \( G_c(x) \) be the probability-generating function for the distribution of finite-length cascades:

\[
G_c(x) \equiv \sum_{c=1}^{S} q_{c} \cdot x^c.
\]  

(21)

The coefficient \( q_{c} \) is the probability that a shock to stock \( s' \)’s fundamentals would set off a cascade of length \( C_s = c \). Let \( G_n(x) \) be the probability-generating function for the number of neighbors:

\[
G_n(x) \equiv \sum_{n=0}^{S-1} p_{n} \cdot x^n.
\]  

(22)

The coefficient \( p_{n} \) is the probability that stock \( s \) has \( n \) neighbors.

Figure 7 shows how these two generating functions are linked. If stock \( s \) has \( n = 1 \) neighbor, \( s' \), then a shock to stock \( s' \)’s fundamentals will set off a cascade of length \( C_s = c \) if a shock the one neighbor will set off a cascade of length \( C_{s'} = c-1 \) excluding stock \( s \). If stock \( s \) has \( n = 2 \) neighbors, \( s' \) and \( s'' \), then a shock to stock \( s \) will set off a cascade of length \( C_s = c \) if shocks to its two neighbors will set off cascades of combined length \( C_{s'} + C_{s''} = c-1 \) excluding stock \( s \). If stock \( s \) has \( n = 3 \) neighbors, then a shock to stock \( s \) will set off a cascade of length \( C_s = c \) if shocks to its three neighbors will set off cascades of combined length \( C_{s'} + C_{s''} + C_{s'''} = c-1 \). And, so on...
\( G_c(x) \) has to satisfy the following internal-consistency condition as the number of stocks gets large, \( S \to \infty \):

\[
G_c(x) = p_0 \cdot x + p_1 \cdot x \cdot G_c(x) + p_2 \cdot x \cdot G_c(x)^2 + p_3 \cdot x \cdot G_c(x)^3 + \cdots \\
= x \cdot \left\{ p_0 \cdot G_c(x)^0 + p_1 \cdot G_c(x)^1 + p_2 \cdot G_c(x)^2 + p_3 \cdot G_c(x)^3 + \cdots \right\} \\
= x \cdot G_n \left( G_c(x) \right).
\]

\( E[C_s] = x \cdot G'_c(x) \bigg|_{x=1} \) gives the expected cascade length when there are no infinitely long cascades, and \( E[N_s] = x \cdot G'_n(x) \bigg|_{x=1} \) gives the expected number of neighbors. Thus:

\[
E[C_s] = x \cdot G'_c(x) \bigg|_{x=1} = 1 + G'_n(1) \cdot G'_c(1) \\
= 1 + E[N_s] \cdot E[C_s].
\]

Rearranging and using the fact that \( E[N_s] = \lambda \) from Lemma 2.1 yields the desired result. \( \square \)

**Proof** (Proposition 2.2b). If \( G_c(x) \) is the probability-generating function for the distribution of finite-length cascades and a fraction \( \theta \) of all cascades are infinitely long, then:

\[
G_c(1) = 1 - \theta.
\]

Using Equation (23c) then implies that the fraction of stocks that would not start an infinitely long rebalancing cascade, \( (1 - \theta) \), is a fixed point of \( G_n(x) \):

\[
1 - \theta = G_c(1) \\
= G_n(G_c(1)) \\
= G_n(1 - \theta).
\]

Since we know from Lemma 2.1 that \( G_n(x) \) is the probability-generating function for the Poisson distribution, we know that \( G_n(x) = e^{\lambda(x-1)} \). Evaluating this formula at \( x = (1 - \theta) \) at taking logs gives the desired result. \( \square \)

**Proof** (Lemma 2.3). The probability-generating function for the number of neighbors that are \( k \)-steps removed from stock \( s \) is given by:

\[
\begin{align*}
1 \text{ step:} & \quad G_n(x) = G_n^{(1)}(x) \\
2 \text{ steps:} & \quad G_n(G_n(x)) = G_n^{(2)}(x) \\
3 \text{ steps:} & \quad G_n(G_n(G_n(x))) = G_n^{(3)}(x) \\
& \quad \vdots \\
k \text{ steps:} & \quad G_n^{(k)}(x).
\end{align*}
\]

So, the average number of \( k \)-step neighbors is given by:

\[
\lambda_k = G'_n(1) \cdot \lambda_{k-1} = \lambda \cdot (\lambda^2/\lambda)^{k-1}.
\]

The average length of the shortest path between two randomly chosen stocks, \( \ell \), is reached when the total number of neighbors at distances \( \leq \ell \) is equal to the total number of stocks in the rebalancing cascade:

\[
\theta \cdot S = 1 + \sum_{k=1}^{\ell} \lambda \ell.
\]

Solving for \( \ell \) when \( (\theta \cdot S) \gg \lambda \) gives the desired result. \( \square \)

**Proof** (Proposition 2.3). To compute the net effect of a long rebalancing cascade starting
with stock \( s_A \) on the demand for stock \( s_Z \), you have to i) find all paths from stock \( s_A \) to stock \( s_Z \) and ii) determine which of these paths cancel each other out. For any 3 paths from stock \( s_A \) to stock \( s_Z \) in a rebalancing cascade, there are 4 possible combinations of outcomes for stock \( s_Z \): \{buy, buy, buy\}, \{buy, buy, sell\}, \{buy, sell, sell\}, and \{sell, sell, sell\}. Whenever the 3 paths create an odd-length cycle, then effects of 2 of the paths will cancel each other out. The combination \{buy, buy, sell\} will just result in a buy order, and the combination \{buy, sell, sell\} will just result in a sell order. Thus, the problem of computing the net effect of a long rebalancing cascade starting with stock \( s_A \) on the demand for stock \( s_Z \) reduces to the problem of finding all odd-length cycles in the rebalancing cascade. There exist polynomial-time algorithms for finding a single cycle of length \( k \) in a random graph (Reed et al., 2004). But, the problem of finding all odd-length cycles (not just a single instance) is at least as hard as the problem of finding the minimal cover of all odd-length cycles. And, the problem of finding the minimal cover of all odd-length cycles reduces to the maximum cut problem (Karp, 1972), which is known to be NP-complete. □
B Tables

Summary Statistics, ETFs

a) Aggregate Data

<table>
<thead>
<tr>
<th></th>
<th>Average</th>
<th>StDev</th>
<th>Min</th>
<th>25%</th>
<th>Median</th>
<th>75%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>#etfs</td>
<td>136.63</td>
<td>43.88</td>
<td>44</td>
<td>99</td>
<td>144</td>
<td>167</td>
<td>220</td>
</tr>
<tr>
<td>\log_{10}(#\text{stocksHeld})</td>
<td>2.49</td>
<td>0.05</td>
<td>2.40</td>
<td>2.46</td>
<td>2.47</td>
<td>2.54</td>
<td>2.60</td>
</tr>
<tr>
<td>\log_{10}(\text{marketValue})</td>
<td>8.20</td>
<td>0.26</td>
<td>7.84</td>
<td>8.02</td>
<td>8.11</td>
<td>8.26</td>
<td>8.84</td>
</tr>
<tr>
<td>#benchmarks</td>
<td>121.17</td>
<td>38.33</td>
<td>38</td>
<td>87</td>
<td>130</td>
<td>151</td>
<td>191</td>
</tr>
<tr>
<td>\langle\text{hasDoubleThreshold}\rangle</td>
<td>0.68</td>
<td>0.02</td>
<td>0.63</td>
<td>0.68</td>
<td>0.68</td>
<td>0.69</td>
<td>0.72</td>
</tr>
</tbody>
</table>

b) Fund-Level Data

<table>
<thead>
<tr>
<th></th>
<th>Average</th>
<th>StDev</th>
<th>Min</th>
<th>25%</th>
<th>Median</th>
<th>75%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>\log_{10}(#\text{stocksHeld})</td>
<td>2.46</td>
<td>0.47</td>
<td>1.00</td>
<td>2.00</td>
<td>2.47</td>
<td>2.78</td>
<td>3.59</td>
</tr>
<tr>
<td>\log_{10}(\text{marketValue})</td>
<td>8.27</td>
<td>1.08</td>
<td>3.48</td>
<td>7.47</td>
<td>8.11</td>
<td>9.12</td>
<td>11.23</td>
</tr>
</tbody>
</table>

Table 1: 

- #etfs: number of ETFs that report portfolio holdings to Morningstar in a given month.
- \log_{10}(\#\text{stocksHeld}): log of the average number of stocks held by each ETF in a given month.
- \log_{10}(\text{marketValue}): log of the average market value of stocks held by each ETF in a given month.
- #benchmarks: number of unique benchmarks used by ETFs in a given month.
- \langle\text{hasDoubleThreshold}\rangle: fraction of ETFs that use a double-thresholded benchmark in a given month.
- \log_{10}(\#\text{stocksHeld}): log of the number of stocks held by each ETF in a given month.
- \log_{10}(\text{marketValue}): log of the market value of the stocks held by each ETF in a given month.

Sparkline plots in Panel (a) depict the time series of monthly observations. Sparkline plots in Panel (b) depict the full distribution of ETF-month observations. Sample: Jan. 2010 to Dec. 2011.
Summary Statistics, Stocks

a) Aggregate Data

<table>
<thead>
<tr>
<th></th>
<th>Average</th>
<th>StDev</th>
<th>Min</th>
<th>25%</th>
<th>Median</th>
<th>75%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>#announcements</td>
<td>11.86</td>
<td>5.62</td>
<td>2</td>
<td>7</td>
<td>12</td>
<td>16</td>
<td>24</td>
</tr>
<tr>
<td>(#unrelatedStocks)</td>
<td>912.09</td>
<td>281.46</td>
<td>369.67</td>
<td>780.08</td>
<td>917.31</td>
<td>970.50</td>
<td>1499.50</td>
</tr>
</tbody>
</table>

b) Stock-Level Data

<table>
<thead>
<tr>
<th></th>
<th>Average</th>
<th>StDev</th>
<th>Min</th>
<th>25%</th>
<th>Median</th>
<th>75%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>log_{10}(etfRebalancing)</td>
<td>1.93</td>
<td>2.03</td>
<td>0.00</td>
<td>0.00</td>
<td>1.93</td>
<td>3.87</td>
<td>7.86</td>
</tr>
<tr>
<td>hasInformedTrading</td>
<td>0.25</td>
<td>0.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>volumeInLast20Min/volume</td>
<td>0.17</td>
<td>0.12</td>
<td>0.05</td>
<td>0.11</td>
<td>0.13</td>
<td>0.20</td>
<td>0.74</td>
</tr>
<tr>
<td>informedVolume/#sharesOutstanding</td>
<td>0.01</td>
<td>0.05</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.81</td>
</tr>
<tr>
<td>#sharesOutstanding</td>
<td>0.01</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.56</td>
</tr>
<tr>
<td>log_{10}(#etfsHoldingStock)</td>
<td>0.82</td>
<td>0.49</td>
<td>0.30</td>
<td>0.30</td>
<td>0.60</td>
<td>1.28</td>
<td>1.97</td>
</tr>
<tr>
<td>amihudIlliquidity</td>
<td>0.04</td>
<td>0.13</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.99</td>
</tr>
<tr>
<td>bidAskSpread</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Table 2: #announcements: number of M&A announcements reported to Thompson Financial in a given quarter. (#unrelatedStocks): average number of unrelated stocks for per M&A announcement in a given quarter. log_{10}(etfRebalancing): log of the ETF-rebalancing volume for each unrelated stock on a given announcement date. hasInformedTrading: indicator variable that is 1 iff a Schedule 13D filer trades an unrelated stock on a given date. volumeInLast20Min/volume: fraction of an unrelated stock’s trading volume that occurs within the final 20 minutes of market close on the announcement day. informedVolume/#sharesOutstanding: informed-trading volume for an unrelated stock on a given announcement date normalized by the number of shares outstanding. ETF trading volume for an unrelated stock on a given announcement date normalized by the number of shares outstanding. log_{10}(#etfsHoldingStock): log of the number of ETFs that hold an unrelated stock at any point during a given month. amihudIlliquidity: Amihud (2002) illiquidity measure for an unrelated stock on a given announcement date. bidAskSpread: average bid-ask spread for an unrelated stock on a given announcement date. Sparkline plots in Panel (a) depict the time series of quarterly observations. Sparkline plots in Panel (b) depict the full distribution of stock-date observations. Sample: Jan. 2010 to Dec. 2011.
Existence of Cascades

Dependent Variable: \( \log(\text{etfRebalancing}_{A \rightarrow Z,t}) \)

<table>
<thead>
<tr>
<th></th>
<th>Full Sample</th>
<th>Hi</th>
<th>Lo</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{afterAnnouncement}_{A,t} )</td>
<td>0.19***</td>
<td>0.13***</td>
<td>0.38***</td>
</tr>
<tr>
<td>( \text{afterAnnouncement}<em>{A,t} \times \text{isTreated}</em>{Z,t} )</td>
<td>0.14***</td>
<td>0.17***</td>
<td>0.13***</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td>( {A, Z, t} )</td>
<td>( {A, Z, t} )</td>
<td>( {A, Z, t} )</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>2.0%</td>
<td>2.0%</td>
<td>2.4%</td>
</tr>
</tbody>
</table>

Table 3: Estimated coefficients for regression specified by Equation (13). Each column reports the results to a separate regression. Data are stock-day level. For each M&A announcement on day \( \tau_A \), we include a 60-day event window, \( t \in [\tau_A - 50, \tau_A + 10) \), of daily observations about each unrelated stock \( s_Z \) on the other side of the market. \( \log(\text{etfRebalancing}_{A \rightarrow Z,t}) \) is the log of the ETF-rebalancing volume for stock \( s_Z \) on day \( t \). \( \text{afterAnnouncement}_{A,t} \) is an indicator variable for whether or not \( t \geq \tau_A \). \( \text{isTreated}_{Z,t} \) is an indicator variable for whether or not stock \( s_Z \) is held by an above-median number of ETFs for the quarter. All regressions include event-\( \tau_A \), stock-\( s_Z \), and year-month fixed effects. Numbers in parentheses are standard errors clustered by stock \( s_Z \). Reads: “In the 10 days after an M&A-related price shock to stock \( s_A \), we estimate that ETF-rebalancing volume for an unrelated stock \( s_Z \) will grow by 14% more when stock \( s_Z \) is held by many ETFs.”
Cascades Generate Noise

Dependent Variable: $\text{hasInformedTrading}_{A\rightarrow Z,t}$

<table>
<thead>
<tr>
<th></th>
<th>Full Sample</th>
<th>endOfDayActivity$_{A,t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Hi</td>
</tr>
<tr>
<td>afterAnnouncement$_{A,t}$</td>
<td>$-0.0001$</td>
<td>$-0.0000$</td>
</tr>
<tr>
<td></td>
<td>$(0.0001)$</td>
<td>$(0.0001)$</td>
</tr>
<tr>
<td>afterAnnouncement$<em>{A,t}$ $\times$ isTreated$</em>{Z,t}$</td>
<td>$0.0004^*$</td>
<td>$0.0006^{**}$</td>
</tr>
<tr>
<td></td>
<td>$(0.0002)$</td>
<td>$(0.0003)$</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td>${A, Z, t}$</td>
<td>${A, Z, t}$</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Table 4: Estimated coefficients for regression specified by Equation (15). Each column reports the results to a separate regression. Data are stock-day level. For each M&A announcement on day $\tau_A$, we include a 60-day event window, $t \in [\tau_A - 50, \tau_A + 10]$, of daily observations about each unrelated stock $s_Z$ on the other side of the market. $\text{hasInformedTrading}_{A\rightarrow Z,t}$ is an indicator variable for whether or not a Schedule 13D filer traded stock $s_Z$ on date $t$. afterAnnouncement$_{A,t}$ is an indicator variable for whether or not $t \geq \tau_A$. isTreated$_{Z,t}$ is an indicator variable for whether or not stock $s_Z$ is held by an above-median number of ETFs for the quarter. All regressions include event-$\tau_A$, stock-$s_Z$, and year-month fixed effects. Numbers in parentheses are standard errors clustered by stock $s_Z$. Reads: “In the 10 days after an M&A-related price to stock $s_A$, we estimate that stock $s_Z$ informed traders are 0.02%pt more likely to trade on days when stock $s_Z$ is held by many ETFs.”
No Evidence of Selection

Dependent Variable: \( \frac{\text{informedVolume}_{s,q}}{\text{#sharesOutstanding}_{s,q}} \)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{\text{etfVolume}<em>{s,q}}{\text{#sharesOutstanding}</em>{s,q}} )</td>
<td>3.42</td>
<td>-22.86**</td>
<td>4.16**</td>
<td>-25.53***</td>
</tr>
<tr>
<td></td>
<td>(2.01)</td>
<td>(10.11)</td>
<td>(2.07)</td>
<td>(10.33)</td>
</tr>
<tr>
<td>log(( \frac{\text{etfsHoldingStock}<em>{s,q}}{\text{#sharesOutstanding}</em>{s,q}} ))</td>
<td>0.01</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(( \frac{\text{etfsHoldingStock}<em>{s,q}}{\text{#sharesOutstanding}</em>{s,q}} )) \times \left( \frac{\text{etfVolume}<em>{s,q}}{\text{#sharesOutstanding}</em>{s,q}} \right)</td>
<td>7.13***</td>
<td>8.45***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.64)</td>
<td>(3.70)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed Effects</td>
<td>( \emptyset )</td>
<td>( \emptyset )</td>
<td>( t )</td>
<td>( t )</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.1%</td>
<td>2.4%</td>
<td>1.3%</td>
<td>3.5%</td>
</tr>
</tbody>
</table>

**Table 5**: Estimated coefficients for regression specified by Equation (17). Each column reports the results of a separate regression. Data are stock-quarter level. \( \frac{\text{informedVolume}_{s,q}}{\text{#sharesOutstanding}_{s,q}} \): informed-trading volume for a stock in a given quarter normalized by the number of shares outstanding. \( \frac{\text{etfVolume}_{s,q}}{\text{#sharesOutstanding}_{s,q}} \): ETF trading volume for a stock in a given quarter normalized by the number of shares outstanding. \( \log(\frac{\text{etfsHoldingStock}_{s,q}}{\text{#sharesOutstanding}_{s,q}}) \): log of the number of ETFs that hold a particular stock at any point during a given quarter. Regressions in columns (3) and (4) include year-quarter fixed effects. Numbers in parentheses are standard errors clustered by stock. Reads: “Informed traders do not trade stock \( s_Z \) different just because it is held by many ETFs.”
Informed Traders Respond to Liquidity

### a) Dependent Variable: \( \text{amihudIlliquidity}_{A\rightarrow Z,t} \)

<table>
<thead>
<tr>
<th></th>
<th>Full Sample</th>
<th>( \text{endOfDayActivity}_{A,t} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{afterAnnouncement}_{A,t} )</td>
<td>0.0018 ( (0.0071) )</td>
<td>0.0056 ( (0.0087) )</td>
</tr>
<tr>
<td>( \text{afterAnnouncement}<em>{A,t} \times \text{isTreated}</em>{Z,t} )</td>
<td>-0.0199** ( (0.0079) )</td>
<td>-0.0386*** ( (0.0107) )</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td>( {A, Z, t} )</td>
<td>( {A, Z, t} )</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.1%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

### b) Dependent Variable: \( \text{bidAskSpread}_{A\rightarrow Z,t} \)

<table>
<thead>
<tr>
<th></th>
<th>Full Sample</th>
<th>( \text{endOfDayActivity}_{A,t} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{afterAnnouncement}_{A,t} )</td>
<td>0.0001*** ( (0.0000) )</td>
<td>0.0002** ( (0.0000) )</td>
</tr>
<tr>
<td>( \text{afterAnnouncement}<em>{A,t} \times \text{isTreated}</em>{Z,t} )</td>
<td>-0.0001** ( (0.0000) )</td>
<td>-0.0003*** ( (0.0001) )</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td>( {A, Z, t} )</td>
<td>( {A, Z, t} )</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.4%</td>
<td>0.2%</td>
</tr>
</tbody>
</table>

**Table 6:** Panel (a): Estimated coefficients for regression specified by Equation (18). Panel (b): Estimated coefficients for regression specified by Equation (19). Each column reports the results to a separate regression. Data are stock-day level. For each M& A announcement on day \( \tau_A \), we include a 60-day event window, \( t \in [\tau_A - 50, \tau_A + 10) \), of daily observations about each unrelated stock \( s_Z \) on the other side of the market. \( \text{amihudIlliquidity}_{s,t} \): Amihud (2002) illiquidity measure for a stock on a given day. \( \text{bidAskSpread}_{s,t} \): average bid-ask spread for a stock on a given day. \( \text{afterAnnouncement}_{A,t} \) is an indicator variable for whether or not \( t \geq \tau_A \). \( \text{isTreated}_{Z,t} \) is an indicator variable for whether or not stock \( s_Z \) is held by an above-median number of ETFs for the quarter. All regressions include event-\( \tau_A \), stock-\( s_Z \), and year-month fixed effects. Numbers in parentheses are standard errors clustered by stock \( s_Z \). Reads: “In the 10 days after an M& A-related price to stock \( s_A \), we estimate that stock \( s_Z \)’s average price impact per dollar traded drops by an additional 1.26% pt when stock \( s_Z \) is held by many ETFs relative to when the same stock is held by only a few ETFs.”