The Sound Of Many Funds Rebalancing*

Alex Chinco[†] and Vyacheslav Fos[‡]

November 5, 2020

Abstract

This paper proposes that computational complexity generates noise. The same asset is often held for completely different reasons by many funds following a wide variety of threshold-based trading rules. Under these conditions, we show it can be computationally infeasible to predict how these various trading rules will interact with one another, turning the net demand from these funds into unpredictable noise. This noise-generating mechanism can operate in a wide range of markets and also predicts how noise volatility will vary across assets. We confirm this prediction empirically using data on exchange-traded funds.

JEL Classification. G00, G02, G14

Keywords. Noise, Thresholds, Complexity, Indexing

*We would like to thank Kerry Back, Nick Barberis, Zahi Ben-David, Bruce Carlin, Maryam Farboodi, Xavier Gabaix, Sam Hartzmark, Ralph Koijen, Pete Kyle, Cecilia Parlatore, Matt Ringgenberg, Gus Sauter, Savi Sundaresan, Chaojun Wang, Brian Weller, and Jeff Wurgler for extremely helpful comments and suggestions. We have also benefited from presentations at Boston College, CalTech, Colorado, Duke, Emory, Illinois, Maryland, Michigan, the SEC, and Yale as well as at the Young-Scholars Finance Consortium, the FINRA Market-Structure Conference, the SFS Calvalcade, the Conference on the Econometrics of Financial Markets, the Helsinki Behavioral Finance Conference, the FRA Conference, the Rodney White Conference, the Adam Smith Conference, the NYU Stern Microstructure Meeting, the CEPR-Imperial-Plato MI3 Conference, and the AFA Annual Meetings.

Current Version: `http://www.alexchinco.com/sound-of-rebalancing.pdf`

[†]The University of Chicago, Booth School of Business; `alex.chinco@chicagobooth.edu`.

[‡]Boston College, Carroll School of Management; `fos@bc.edu`.

1 Introduction

In a market without noise, a trader who discovers that an asset is under-priced cannot take advantage of this discovery (Aumann, 1976; Milgrom and Stokey, 1982). The moment he tries to buy a share, other traders will immediately realize he must have uncovered good news. No one will agree to sell him any shares at the old price (Grossman, 1976).

Noise pulls the rug out from under this no-trade theorem. In a market with noise, there are always some shares of the asset being bought or sold for erratic non-fundamental reasons. So, if the trader tries to buy a share, other traders will not immediately realize that he has uncovered some good news. His buy order might just be some more random noise. This cover story allows the trader to profit from his discovery. It is this plausible alibi that "makes financial markets possible (Black, 1985)."

But, where does noise come from? Previous researchers have pointed to several different mechanisms. Early papers suggested that noise comes from random supply shocks (Grossman and Stiglitz, 1980; Hellwig, 1980; Verrecchia, 1982; Admati, 1985) or from liquidity traders with random demand for cash (Glosten and Milgrom, 1985). There is a large literature on noise traders (Shiller, 1984; De Long et al., 1990; Shleifer and Summers, 1990) whose random trades stem from irrational beliefs. Other papers model noise as the result of agents' need to hedge random endowment shocks (Dow and Gorton, 1997).

This paper proposes an alternative noise-generating mechanism: computational complexity. In modern financial markets, the same asset is often held for completely different reasons by funds following a wide variety of threshold-based trading rules. We show that, under these conditions, it is computationally infeasible to predict how the various trading rules will interact with one another. As a result, the net demand coming from an interacting mass of funds will appear random even if market participants are fully rational and the individual trading rules involved are simple.

To show how computational complexity generates noise, we study a theoretical model motivated by three common features of modern financial markets. First, we assume there are a large number of funds which follow a wide variety of different trading rules. This statement is self evident, and it applies just as well to hedge funds as to mutual funds, pension funds, algorithmic traders, and index funds. Second, because markets are populated by such a large heterogeneous collection of funds, we assume that the same asset is often held by different funds for completely different reasons. For example, one fund's value stock might be another fund's low-volatility stock. Third, we assume that many funds use threshold-based trading rules. For instance, the PowerShares S&P 500 Low-Volatility ETF [SPLV] tracks a benchmark consisting of the 100 lowest volatility S&P 500 stocks. This is a threshold-based trading rule because an arbitrarily small change in a stock's volatility can move it from 101st to 100th place on the low-volatility leaderboard. When this happens, SPLV will exit its position in that stock and build a new position in another, affecting each asset in equal-but-opposite ways.

When there are so many different funds with overlapping holdings using threshold-based trading rules, a small change in asset A's price can cause one fund to buy asset A and sell asset B, which can then cause a second fund using a different threshold to sell asset B and buy asset C, which can then cause... Under these widely observed conditions, we prove it is possible to determine *if* an unrelated asset Z will be affected by one of these rebalancing cascades starting with a shock to stock A. But, the problem of determining *how* asset Z will be affected (buy? or sell?) is computationally intractable.

Formally, we show this is an NP-hard problem. In a large market, it is computationally infeasible to predict the demand coming from a rebalancing cascade unless, at minimum, a polynomial-time algorithm can be found that solves every NP problem. No such algorithm is known, and it is widely believed that no such algorithm exists. But, proving this fact remains "the central unsolved problem of theoretical computer science (Aaronson, 2013)." The sign of the resulting demand shock may as well be thought of as a coin flip—i.e., it may as well be noise. This remains true even if agents are fully rational and each fund involved in the cascade follows a simple deterministic trading rule.

What does this theoretical analysis reveal? The key insight is not that financial markets can be computationally complex in a few special cases or that the resulting shocks might appear random to most unsophisticated agents. Rather, the model shows how three common market features can combine to produce demand shocks which all agents find computationally intractable to predict. It is this combination—ubiquitous features, universal intractability—that is new and noteworthy. It is this combination that suggests computational complexity is an important noise-generating mechanism in financial markets.

There are three reasons why the computational complexity of rebalancing cascades is an especially useful noise-generating mechanism to know about. First, rebalancing cascades can generate noise in a broad range of markets and timescales. By contrast, although noise traders produce random shocks at human timescales (Barber and Odean, 2000), it is unlikely that they also do so at the much faster timescales where most trading activity now takes place (Pagnotta and Philippon, 2018). Second, this noise-generating mechanism offers another reason why the rise of sophisticated investors may not "ultimately help to make markets more efficient (Stein, 2009)." If markets with more sophisticated investors also contain a more diverse collection of funds, then more investor sophistication will be accompanied by more noise from rebalancing cascades.

Third, and most importantly for researchers, this mechanism predicts how noise volatility will vary across assets, which offers a new way to examine economic models with noise-dependent predictions. There are many such models. For example, illiquidity is a function of both fundamental volatility and noise volatility in Kyle (1985). Likewise, Hellwig and Veldkamp (2009) show how small changes in noise volatility can produce large changes in prices, allocations, and welfare when there are strategic complementarities between information acquisition and portfolio choice.

We test this prediction using data from ETF Global on the daily holdings of ETFs from January 2011 to December 2017. This data allows us to study the network of stocks linked by potential ETF rebalancing decisions. Stock A is positively linked to stock B in this network if a drop in its fundamentals would cause some ETF to replace its position in stock A with a position in stock B. Conversely, stock A would be negatively linked to stock B if an increase in the fundamentals of stock A would cause some ETF to dump its holdings of stock B and build a position in stock A. In either case, we say that stocks A and B are on the cusp of a rebalancing threshold. The ETF rebalancing network is the network of stocks connected by these potential ETF rebalancing decisions. Just as predicted by the model, we empirically verify that stocks on the cusp of more ETF rebalancing thresholds experience more demand noise coming from ETF rebalancing cascades—a sequence of ETF rebalancing decisions stemming from an initial shock to stock A.

Before going any further it is important to clarify two things about our data. First, we are netting-out changes in ETF holdings due to creations and redemptions since these trades are executed as in-kind transfers for tax reasons (Madhavan, 2016; Ben-David et al., 2017). Second, we are also restricting our data to only include ETFs that rebalance more than once a quarter. While these ETFs tend to be smaller, their rebalancing still matters since they tend to trade during the 20-to-30 minutes before market close.[1]

In our theoretical model, we study rebalancing cascades that stem from an initial shock to asset A, which is publicly observed. So, in our empirical analysis, we have to make a decision about which initial shocks to use. To this end, we focus on ETF rebalancing activity stemming from M&A announcements, referring to the target of the M&A announcement

[1]Wall Street Journal. 5/27/2015. *Stock-Market Traders Pile In at the Close*.

as stock A. M&A deals are a natural choice because "a profusion of event studies has demonstrated that mergers seem to create shareholder value, with most of the gains accruing to the target company (Andrade et al., 2001)."

To make sure we are analyzing ETF trading activity coming indirectly through rebalancing cascades and not directly from the original M&A-announcement shocks themselves, we study the ETF rebalancing activity of stocks that are unrelated to each M&A target, stock A. We refer to these unrelated stocks as "stock Zs". For stock A and stock Z to be unrelated, they must be twice removed in the network of ETF holdings at the time of the M&A announcement. Stock Z cannot have been recently held by any ETF that also recently held stock A. Likewise, if stock A and stock B were both recently held by the same ETF, then stock Z also cannot have been recently held by any ETF that recently held stock B. The chain of ETF rebalancing decisions connecting stock A and stock Z must be $A \to B \to C \to Z$ or longer. Because there are smart-beta ETFs tracking things like large-cap, value, and industry, this criteria implies stock A and stock Z have dissimilar factor exposures and firm characteristics.

We split the set of stock Zs that are unrelated to the target of each M&A announcement, stock A, into two subsets: those that are on the cusp of rebalancing for an above-median number of ETFs, and those that are not. After verifying that there are no differences in the pre-trends in the days leading up to the M&A announcement between these two groups of stocks, we first show that ETF rebalancing volume is 2.06% higher for a stock Z in the five days immediately after an M&A announcement when it sits on the cusp of an above-median number of ETF rebalancing thresholds. Then, we show that this increase is no more likely to be made up of buy orders than of sell orders. Our identification comes from the precise timing of the announcements and the fact that the same stock Z can be above-median relative to one stock A's announcement while below-median relative to another. In short, you can predict *if* stock Z will be affected by an ETF rebalancing cascade stemming from an unrelated stock A but not *how* stock Z will be affected.

In canonical asset-pricing models such as Grossman and Stiglitz (1980) and Kyle (1985), noise shocks are fundamentally unpredictable random variables that affect equilibrium asset prices but are unrelated to firm fundamentals. Moreover, investors in these models know the distribution from which these unpredictable shocks are drawn; they just do not know the realized value of the noise shock. The demand shocks coming from long rebalancing cascades fit these criteria. We provide a theoretical guarantee that the direction of a demand shock to stock Z coming from a long rebalancing cascade is as random as the outcome of a coin flip. Moreover, this demand shock has nothing to do with stock Z's fundamentals since the rebalancing cascade is the result of an initial shock to an unrelated stock A. Thus, because it is possible to predict if a stock Z will be hit by an ETF rebalancing cascade, investors can predict the volatility of the resulting demand shocks. In short, our analysis gives a constructive theory of where the noise shocks, which play such a pivotal role in our asset-pricing models, come from. It explains where these shocks come from, why they are fundamentally unpredictable, and how investors can predict demand-noise volatility.

1.1 Related Literature

This paper borrows from and builds on three main strands of literature.

Noise. Noise plays a central role in financial markets via its impact on liquidity (Grossman and Stiglitz, 1980; Kyle, 1985), limits to arbitrage (Shleifer and Summers, 1990; Shleifer and Vishny, 1997; Gromb and Vayanos, 2010), and information aggregation (Hellwig, 1980; Admati, 1985). The key contribution of this paper is to propose a noise-generating mechanism that does not rely on external randomness. What's more, the mechanism proposed in this paper predicts which assets should have more/less noise.

Complexity. Numerous other researchers have studied complexity and chaos in financial markets (e.g., Baumol and Benhabib, 1989; Frank and Stengos, 1989; Scheinkman and LeBaron, 1989; Hsieh, 1991; Brock and Hommes, 1997; Rosser, 1999). We are not the first. This paper was clearly inspired by this ambitious and exciting earlier work. But, we are trying to move this line of research in a new direction.

Existing research uses complexity to show how economic phenomena, such as financial crises, are fundamentally unpredictable. Most of the papers cited above were published *en masse* following the 1987 crash. This paper is pointing out a constructive application of these so-called *'no go'* theorems. Computational complexity can generate the noise which makes financial markets possible and it can also predict where this noise will be the loudest, giving new ways of testing existing economic theories.

Indexing. Finally, our empirical application connects to research on index-linked investing (Wurgler, 2010), which can be classified into two subgroups. The first group of papers studies how index inclusion affects the underlying stocks in a predictable way. For instance, papers such as Chang et al. (2014), Bessembinder (2015), Shum et al. (2015), Bai et al. (2015), and Ivanov and Lenkey (2018) all study the predictable effects of index rebalancing decisions on asset A. For still more examples, see Ben-David et al. (2017), Bessembinder et al. (2016), Brown et al. (2016), and Israeli et al. (2017).

The second group studies how asset A's inclusion will affect its relationship with some other asset B in a predictable fashion. For instance, Barberis et al. (2005) shows that a stock's beta with the S&P 500 jumps sharply after inclusion. See also Greenwood and Thesmar (2011), Vayanos and Woolley (2013), and Anton and Polk (2014). By contrast, this paper focuses on the unpredictable consequences of asset A's index inclusion, not for asset A or asset B, but for a completely unrelated asset Z.

2 Theoretical Model

This section presents a theoretical model where, although it is possible to determine *if* an asset will be affected by a rebalancing cascade, it is nevertheless computationally infeasible to determine *how* that asset will be affected (buy? or sell?). So, market participants will think about the demand shocks coming from these rebalancing cascades as effectively random. The key insight delivered by our theoretical analysis is that common market features can combine to produce demand shocks which are provably intractable to predict for all market participants.

2.1 Market Structure

Here is how we model funds transmitting an initial shock from asset A to asset B, and then from asset B to asset C, and then from asset C to asset D, and so on via their rebalancing rules. These rebalancing rules are going to be extremely simple. This is precisely the point. One of the goals of the model is to show how complexity can generate noise even if individual agents are following extremely simple decision rules. Using a more realistic set of rules will make prediction even harder.

Network. Imagine a market where rebalancing rules define a network over a set of assets $\mathsf{S} = \{1, 2, \ldots, S\}$. There is an edge from asset s to asset s', not if they are both currently held by the same fund, but rather if a shock to asset s would cause a fund to swap its position in asset s for a new position in asset s'. If a positive shock to asset s would cause some fund to sell asset s' and buy asset s, then asset s' is a negative neighbor to asset s, $\mathsf{N}_s^- = \{s' \in \mathsf{S} \,|\, \text{positive shock to } s \Rightarrow \text{negative shock to } s'\}$. If a negative shock to asset s would cause some fund to buy asset s' and sell asset s, then asset s' is a positive neighbor of asset s, $\mathsf{N}_s^+ = \{s' \in \mathsf{S} \,|\, \text{negative shock to } s \Rightarrow \text{positive shock to } s'\}$. A market's structure is the set of neighbors for each asset, $\mathsf{M} = \{(\mathsf{N}_1^+ \| \mathsf{N}_1^-), \ldots, (\mathsf{N}_S^+ \| \mathsf{N}_S^-)\}$.

A shock to asset s could be an information event, such as an announcement that the stock is the target of an M&A deal. A shock could also be a change in the market characteristics of asset s, such as an increase in its return volatility or its trading volume. Our empirical analysis studies how an initial shock coming from an information event propagates through the rebalancing network due to changes in the market characteristics of each asset. Because of the way we have defined the rebalancing network (a link from asset s to asset s' exists if a shock to asset s would cause a fund to execute a trade involving assets s and s'), we will talk about shocks and trades interchangeably throughout the rest of the paper.

Distortion. We want to analyze how this rebalancing network propagates shocks through the market in discrete rounds indexed by $t = 0, 1, 2, \ldots$ For this to happen, rebalancing decisions must have the potential to distort asset characteristics. If one fund decides to sell asset s, then this decision must have the potential to change asset s in a way that causes a second fund to rebalance, as well. This assumption is consistent both with trader descriptions and current academic research (Ben-David et al., 2017).

We embed this assumption in our model by using a single variable, $x_{s,t}$, to keep track of both rebalancing decisions and changes in asset characteristics:

$$x_{s,t} \in \{-1,\ 0,\ 1\} \qquad\qquad \Delta x_{s,t} = x_{s,t} - x_{s,t-1} \tag{1}$$

Note that, because of the structure of the rebalancing cascades we model, changes in asset characteristics will only take on values $\Delta x_{s,t} \in \{-1, 0, 1\}$ as will become clear shortly. If $(x_{s,t}, \Delta x_{s,t}) = (1, 1)$, then asset s has realized a positive shock because some fund built a new position in that asset. If $(x_{s,t}, \Delta x_{s,t}) = (-1,\ -1)$, then the opposite outcome has taken place. Note that Duffie et al. (2005) use a model with a similar ternary structure. To emphasize that rebalancing decisions affect more than just an asset's price, we refer to changes in *characteristics* rather than prices. For example, if a large-cap equity ETF decides to buy a stock, then this additional buying pressure might increase the stock's volatility enough to force a second low-vol equity ETF to exit its position in the stock.

There are cascade-like phenomena without alternation. For example, think about bank runs. However, rebalancing decisions necessarily involve alternation. They require a fund to replace its position in one stock with a position in another. Section 2.5 describes in more detail how alternation plays a key role in the making it computationally intractable to predict how a long rebalancing cascade will affect the demand for a given stock. A key insight in this paper is that a seemingly innocuous fact about each individual rebalancing decision—that it delivers a positive shock to one asset and a negative shock to another—can make a big difference when it comes to predicting the outcome of long rebalancing cascades.

Real world rebalancing decisions can certainly be more complicated than the ones the model in this paper. However, many funds do make simple one-for-one decisions. Our theoretical analysis shows how rebalancing cascades involving just these sorts of simple decisions—i.e., rebalancing decisions you might expect would be the easiest to predict the outcomes of—have the capacity to generate unpredictable demand shocks.

Propagation. Because we want to illustrate how computational complexity can generate seemingly random demand shocks even in the absence of any random behavior on the part of individual investors, we model how rebalancing decisions propagate shocks through the market as a mechanical deterministic three-step process. First, Step 1 involves identifying all assets that will be affected at time $(t+1)$ by rebalancing decisions at time t:

$$\mathsf{Out}^{+}_{s,t} = \begin{cases} \{s' \in \mathsf{N}^{+}_{s} \mid s \notin \mathsf{Out}^{-}_{s',t-1}\} & \text{if } (x_{s,t}, \Delta x_{s,t}) = (-1,\ -1) \\ \emptyset & \text{otherwise} \end{cases} \tag{2a}$$

$$\mathsf{Out}^{-}_{s,t} = \begin{cases} \{s' \in \mathsf{N}^{-}_{s} \mid s \notin \mathsf{Out}^{+}_{s',t-1}\} & \text{if } (x_{s,t}, \Delta x_{s,t}) = (1, 1) \\ \emptyset & \text{otherwise} \end{cases} \tag{2b}$$

$\mathsf{Out}^{-}_{s,t}$ is the set of assets that will be negatively affected at time $(t+1)$ by some fund's decision to buy asset s at time t. Likewise, $\mathsf{Out}^{+}_{s,t}$ is the set of assets that will be positively affected at time $(t+1)$ by some fund's decision to sell asset s at time t. The restrictions that $s \notin \mathsf{Out}^{-}_{s',t-1}$ and $s \notin \mathsf{Out}^{+}_{s',t-1}$ respectively ensure that a shock does not just bounce back and forth between assets s and s' in perpetuity.

Next, Step 2 involves identifying all the ways that each asset $s \in \mathsf{S}$ will be affected at time $(t+1)$ by this collection of outgoing links due to decisions made at time t:

$$\mathsf{In}^{+}_{s,t+1} = \{s' \in \mathsf{S} \mid s \in \mathsf{Out}^{+}_{s',t}\} \tag{3a}$$

$$\mathsf{In}^{-}_{s,t+1} = \{s' \in \mathsf{S} \mid s \in \mathsf{Out}^{-}_{s',t}\} \tag{3b}$$

Positive incoming links for asset s correspond to situations where a fund sold asset s' at time t, and this selling pressure then forced a second fund following a different trading rule to buy asset s at time $(t+1)$. Negative incoming links for asset s correspond to the same sequence of events with opposite signs.

Finally, Step 3 involves calculating how the net effect of this collection of incoming links will distort the characteristics of each asset at time $(t+1)$:

$$u_{s,t+1} = 1_{\{|\mathsf{In}^{+}_{s,t+1}|>|\mathsf{In}^{-}_{s,t+1}|\}} - 1_{\{|\mathsf{In}^{+}_{s,t+1}|<|\mathsf{In}^{-}_{s,t+1}|\}} \tag{4a}$$

$$x_{s,t+1} = \mathrm{Sign}[x_{s,t} + u_{s,t+1}] \tag{4b}$$

In the equation above, $\mathrm{Sign}[y] = y/|y|$. This updating rule says that, if more funds decided to buy asset s than sell asset s at time $(t+1)$, then the asset will realize a positive shock; if more funds decided to sell, then asset s will realize a negative shock.

Cascades. Rebalancing cascades start at $t = 0$ with all asset characteristics set to

$$(x_{s,0}, \Delta x_{s,0}) = (0,0) \tag{5}$$

Then, at $t = 1$, nature selects an ϵ-small set, A, to receive an initial positive shock:

$$(x_{s,1}, \Delta x_{s,1}) = (1,1) \qquad\qquad \text{for each } s \in \mathsf{A} \tag{6}$$

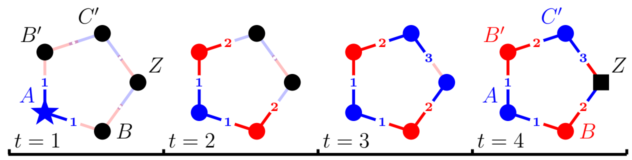
The positive-initial-shock convention is without loss of generality. We assume that everyone knows the identity of the assets in A. We say that A is ϵ-small if there is a positive constant $\epsilon > 0$ such that $|\mathsf{A}| < \epsilon \cdot S$ as $S \to \infty$.

Our goal is to outline how a simple deterministic process can produce unpredictable demand shocks. So, following the initial positive shock, a rebalancing cascade is just the iteration of the 3-step updating procedure outlined above until a time limit $T \in \{1, 2, \ldots\}$ has been reached. A rebalancing cascade's effect on asset Z, $\text{Effect}_{\mathsf{M},T}(\mathsf{A}, Z)$, is the Zth element of the output from $\text{Cascade}_{\mathsf{M},T}(\mathsf{A})$:

function $\text{Cascade}_{\mathsf{M},T}(\mathsf{A})$**:**
    $t \leftarrow 0$
    for all $(s \in \mathsf{A})$**:**
        $(x_s, \Delta x_s) \leftarrow (1,1)$
    while $(t \leq T)$**:**
        for all $(s \in \mathsf{S})$**:**
STEP 1: $(\mathsf{Out}_s^+, \mathsf{Out}_s^-) \leftarrow \text{Update}[(\mathsf{Out}_s^+, \mathsf{Out}_s^-) | (x_s, \Delta x_s)]$
        for all $(s \in \mathsf{S})$**:**
STEP 2: $(\mathsf{In}_s^+, \mathsf{In}_s^-) \leftarrow \text{Update}[(\mathsf{In}_s^+, \mathsf{In}_s^-)]$
STEP 3: $(x_s, \Delta x_s) \leftarrow \text{Update}[(x_s, \Delta x_s)]$
        $t \leftarrow t + 1$
    return $[\, x_1 \;\, x_2 \;\cdots\; x_S \,]$

Notice how the description of a rebalancing cascade suggests a second interpretation for the symbol M. The symbol M does not just represent a description of the rebalancing rules. It is also a description of a machine that computes the effect of the resulting cascades.

An Example. Figure 1 shows an example of a rebalancing cascade involving 5 assets that starts with a positive shock to asset A. Notice that the cascade involves both alternating



	N_s^+	N_s^-	$t=1$	$t=2$	$t=3$	$t=4$, $\text{Cascade}_{\mathsf{M},4}(\mathsf{A})$
A		$\Vert B, B'$	$+1$	$+1$	$+1$	$+1$
B	$A, Z \Vert$		0	-1	-1	-1
B'	$A, C' \Vert$		0	-1	-1	-1
C'		$\Vert B', Z$	0	0	$+1$	$+1$
Z	$C' \Vert$	B	0	0	$+1$	$0 = \text{Effect}_{\mathsf{M},4}(\mathsf{A}, Z)$

Figure 1. An Example. A rebalancing cascade involving 5 assets that starts with a positive shock to asset A, $\mathsf{A} = \{A\}$. Grey box depicts market structure, M. Columns to the right depict state of each asset, $x_{s,t}$, at times $t = 1, 2, 3, 4$. Diagrams above depict propagation of shock through the market. Dots denote assets. Red dot is $x_{s,t} = -1$; blue dot is $x_{s,t} = +1$; black dot is $x_{s,t} = 0$. Dashed box reports result of rebalancing cascade at time $t = 4$, $\text{Cascade}_{\mathsf{M},4}(\mathsf{A})$. Cascade has positive effect on asset Z in round $t = 3$, $\text{Effect}_{\mathsf{M},3}(\mathsf{A}, Z) = +1$. But, in round $t = 4$, effect reverts to $\text{Effect}_{\mathsf{M},4}(\mathsf{A}, Z) = 0$.

sequences of buy and sell orders as well as cancellations. By alternation, we mean that when stock s receives a shock at time t, its neighbors receive a shock of the opposite sign at time $(t + 1)$. Here is what we mean by cancellations. At time $t = 3$, the cascade delivers a positive shock to asset Z, $\text{Effect}_{\mathsf{M},3}(\{A\}, Z) = +1$. But then, at time $t = 4$, a second branch of the cascade hits asset Z, canceling out the effect of the first shock, $\text{Effect}_{\mathsf{M},4}(\{A\}, Z) = 0$. This example highlights the two questions we want to ask about rebalancing cascades in the following two subsections. First, is there any way for a rebalancing cascade that starts at asset A to eventually affect asset Z? Second, suppose there is. What will be the net effect of the rebalancing cascade on asset Z?

2.2 'If?' Problem

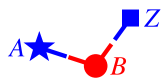
How hard is it to figure out whether a rebalancing cascade triggered by an initial shock to asset A might eventually affect the demand for asset Z? It turns out that the answer to this question is: *'Not very hard.'* We now explore why.

Decision Problem. Figuring out whether a rebalancing cascade starting with asset A might affect asset Z means finding at least one path connecting A to Z in the rebalancing network. We define a K-path connecting asset A to asset Z as a sequence of K assets

$\{s_1, \ldots, s_K\}$ such that the first asset is asset A, the last asset is asset Z, and for each asset $k \in \{2, \ldots, K-1\}$ we have that

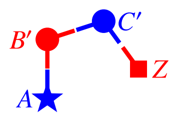
$$s_k \in \begin{cases} \mathsf{N}^{+}_{s_{k-1}} & k \text{ odd} \\ \mathsf{N}^{-}_{s_{k-1}} & k \text{ even} \end{cases} \qquad \text{for all } k \in \{2, \ldots, K\} \tag{7}$$

In Figure 1, there are two different paths from A to Z. One travels from A to B to Z:



$$\{\underset{\textbf{Asset } \boldsymbol{A} \,\rightarrow}{(\emptyset \| \{B\})}, \underset{\textbf{Asset } \boldsymbol{B} \,\rightarrow}{(\{A,Z\} \| \emptyset)}, \underset{\textbf{Asset } \boldsymbol{Z}}{(\emptyset \| \{B\})}\} \tag{8}$$

The other travels from A to B' to C' to Z:



$$\{\underset{\textbf{Asset } \boldsymbol{A} \,\rightarrow}{(\emptyset \| \{B'\})}, \underset{\textbf{Asset } \boldsymbol{B'} \,\rightarrow}{(\{A,C'\} \| \emptyset)}, \underset{\textbf{Asset } \boldsymbol{C'} \,\rightarrow}{(\emptyset \| \{B',Z\})}, \underset{\textbf{Asset } \boldsymbol{Z}}{(\{C'\} \| \emptyset)}\} \tag{9}$$

If such a path exists, it is possible a rebalancing cascade triggered by an initial shock to asset A might affect the demand for asset Z.

Below we formally define the 'If?' decision problem.

Problem 2.2a (If)**.**

- *Instance: An asset Z, market structure* M*, time* $T \geq 1$*, and subset of assets* $\hat{\mathsf{S}} \subseteq \mathsf{S}$.
- *Question: For each asset $s \in \hat{\mathsf{S}}$, is there a K-path connecting s to asset Z for $K \leq T$?*

The set If denotes the collection of instances where the answer is *'Yes'*. Solving the 'If?' problem means deciding whether $(Z, \mathsf{M}, T, \hat{\mathsf{S}}) \in \mathsf{If}$. So, when $(Z, \mathsf{M}, T, \hat{\mathsf{S}}) \in \mathsf{If}$, there is at least one K-path connecting each asset $s \in \hat{\mathsf{S}}$ to asset Z in $K \leq T$ steps.

If Complexity. Problems with polynomial-time solutions are "tractable problems" that "can be solved in a reasonable amount of time (Moore and Mertens, 2011)." The proposition below shows that If can be solved in polynomial time. It is easy to determine which asset As have the potential to trigger a rebalancing cascade affecting asset Z.

Proposition 2.2a (If Complexity)**.** If *can be solved in polynomial time.*

We say that $\mathrm{f}(y) = \mathrm{O}[\mathrm{g}(y)]$ if there exists an $\alpha > 0$ and a $y_0 > 0$ such that $|\mathrm{f}(y)| \leq \alpha \cdot |\mathrm{g}(y)|$ for all $y \geq y_0$. We say that $\mathrm{f}(y) = \mathrm{Poly}[y]$ if there exists some $\beta > 0$ such that $\mathrm{f}(y) = \mathrm{O}[y^{\beta}]$. The size of an instance of If is governed by the number of assets in the market, S. So, a polynomial-time solution for If is an algorithm that decides whether $(Z, \mathsf{M}, T, \hat{\mathsf{S}}) \in \mathsf{If}$ in $\mathrm{Poly}[S]$ steps for every possible choice of $(Z, \mathsf{M}, T, \hat{\mathsf{S}})$—i.e., computational-complexity results typically provide bounds on the time needed to solve worst-case instances. For further details, see the numerical simulations in Section 2.4 (Figure 4) as well as our discussion of time complexity in the Internet Appendix.

Predicting If. The tractability of the If problem also means that a researcher can make useful predictions about the size of $\hat{\mathsf{S}}$, the set of stocks connected to a given asset Z. To illustrate, suppose that for any pair of assets $(s, s') \in \mathsf{S}^2$, asset s' is chosen as a positive neighbor to asset s independently with probability κ/S where $\kappa > 0$ is some $\mathrm{O}[\log S]$ function. Under these assumptions, the number of positive neighbors for each asset, $N_s^+ = |\mathsf{N}_s^+|$, obeys a Poisson distribution as $S \to \infty$ (Erdos and Rényi, 1960)

$$N_s^+ \sim \text{Poisson}(\kappa, S) \tag{10}$$

which implies the typical asset has $\mathrm{E}[N_s^+] = \kappa$ positive neighbors. Thus, if $\kappa \approx 0$, the market will be fragmented. Most assets will have no neighbors. Whereas, if $\kappa \approx \log S$, the market will be densely connected with each asset on the cusp of rebalancing for many different funds. The fact that there are "now more indexes than stocks"[2] suggests $\kappa \gg 1$. Our data suggest that $\kappa \approx \frac{1}{2} \cdot \log S$ in the ETF rebalancing network. There are roughly $S = 1{,}500$ stocks in our sample, and Panel B of Table 1 reports that the average stock is on the cusp of rebalancing for 3.4 different ETFs, which is roughly $\frac{1}{2} \cdot \log(1{,}500) \approx 3.65$.

The proposition below shows that it is easy to predict how many assets are connected to asset Z just by counting the number of neighbors for asset Z.

Proposition 2.2b (Predicting If)**.** *If* M *is a market structure generated using* $\kappa > 1$ *and*

$$\hat{\mathsf{S}}_{\max}(Z, \mathsf{M}, T) = \max_{\hat{\mathsf{S}} \in 2^{\mathsf{S}}} \{\, |\hat{\mathsf{S}}| \ s.t.\ (Z, \mathsf{M}, T, \hat{\mathsf{S}}) \in \mathsf{If} \,\} \tag{11}$$

denotes the number of assets with a K-path to asset Z for some $K \leq T$*, then* $\mathrm{E}[\hat{\mathsf{S}}_{\max}(Z, \mathsf{M}, T)]$ *is increasing in the total number of neighbors to asset Z.*

Put differently, assets with more neighbors are more likely to be affected by a rebalancing cascade. What's more, you can infer this property without having to trace out each individual path that a rebalancing cascade might take. All you have to do is count the number of immediate neighbors to asset Z. We use this fact in our empirical analysis.[3]

[2]Bloomberg. 5/12/2017. *There Are Now More Indexes Than Stocks*.

[3]We want to emphasize that our main results are not about the emergence of a giant component in a random graph when the connectivity parameter crosses $\kappa = 1$. In fact, Proposition 2.2b takes the existence of the giant connected component as given when it assumes $\kappa > 1$. This is a reasonable assumption to make in our empirical analysis since modern financial markets contain more indexes than stocks. As discussed above, the typical stock in our data set has 3.4 neighbors in the ETF rebalancing network. Our main point is that the interactions between a large number of heterogeneous rebalancing rules in this giant component can create seemingly random demand shocks.

2.3 'How?' Problem

Although it is easy to predict *if* some asset Z is likely to be affected by a rebalancing cascade starting with asset A, predicting *how* this asset will be affected turns out to be computationally intractable. Let's now examine why this is the case.

Some Intuition. What does it mean to say that *if* is an easier question than *how*? To build some intuition, let's start by looking at Figure 2. Each row depicts a single market with $S = 25$ assets. Here is the exercise we have in mind. First, examine the left panel, which depicts the rebalancing rules that define each market. Then, ask yourself two questions: #1) *'Will asset Z, which is denoted by the large black square with a question mark in it, be affected by a rebalancing cascade that starts at asset A, which is denoted by the large blue star?'* and #2) *'If so, how exactly will asset Z be affected (buy vs. sell)?'*

On one hand, you can immediately *see* how easy it is to answer question #1. The middle panels show there is a path connecting asset A to asset Z in markets M_2, M_3, and M_4 (rows two through four) but not in market M_1 (row one). So, asset Z might be affected by a rebalancing cascade starting with asset A in M_2, M_3, and M_4 but not in M_1. Answering this first question gives you a sense of what a polynomial-time solution looks like. All you have to do is find a break in the chain.

On the other hand, you can also immediately *see* how hard it is to answer question #2. The middle column shows that markets M_2, M_3, and M_4 (rows two through four) all have individual paths connecting asset A to asset Z that end with a positive demand shock for asset Z. But, the right column shows that the net effect of the entire rebalancing cascade in each of these markets only agrees with this naïve prediction for market M_2 (second row). Even though there are individual paths from asset A to asset Z that would result in a positive demand shock for asset Z in markets M_3 and M_4, neither rebalancing cascade results in a positive demand shock for asset Z on net due to cancellations. There is simply no way to guess how a rebalancing cascade will affect asset Z by examining the set of rebalancing rules involved, even though these rules are completely deterministic.

Decision Problem. Below is the formal definition of the 'How?' decision problem.

Problem 2.3a (How)**.**

- *Instance: An asset Z, market structure* M*, time* $T = \text{Poly}[S]$*, positive constant* $\epsilon > 0$*, and power set* $\hat{\mathsf{A}} \subseteq 2^{\mathsf{S}}$ *of all* ϵ*-small sets* $\mathsf{A} \subseteq \mathsf{S}$.
- *Question: Is there some* $\mathsf{A} \in \hat{\mathsf{A}}$ *such that* $\text{Effect}_{\mathsf{M},T}(\mathsf{A}, Z) \neq +1$*?*

The set How denotes the collection of instances where the answer is *'Yes'*. Here is what this

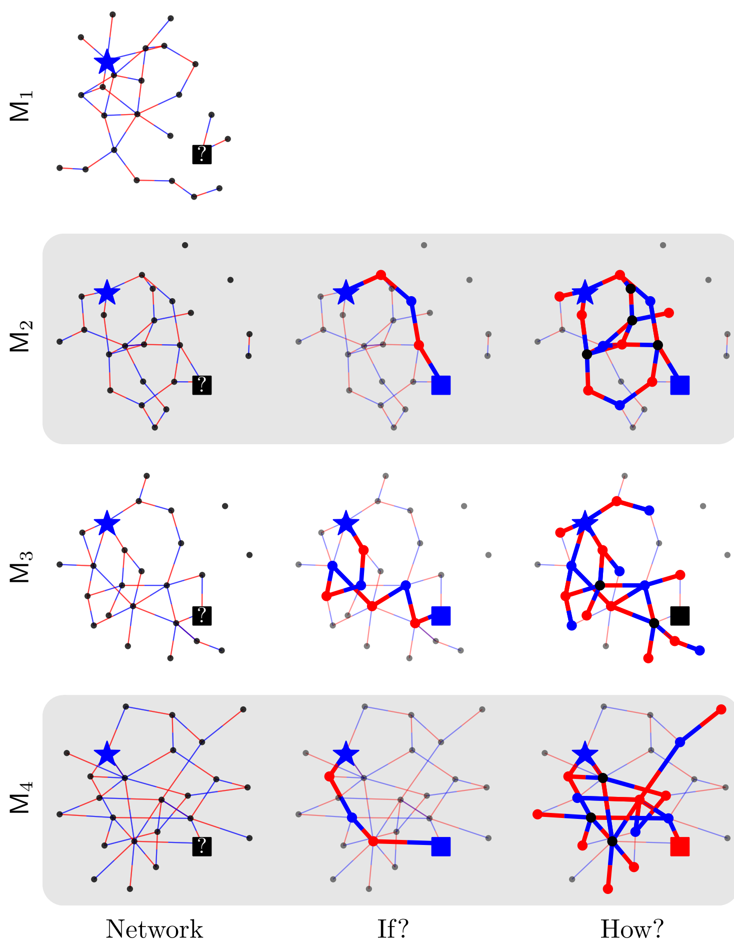


Figure 2. Some Intuition. Each row contains 3 panels and depicts simulated results for a single market with $S = 25$ assets—i.e., one market structure per row. Nodes are assets. Node color denotes effect of rebalancing cascade: blue=positive, red=negative, black=no effect. Star: asset A. Square: asset Z. Edges denote rebalancing rules. Blue(s)-to-red(s'): asset s' is negative neighbor to asset s. Red(s)-to-blue(s'): asset s' is positive neighbor to asset s. Asset A and asset Z are in same position in all panels. Network: rebalancing rules. If?: Path connecting asset A to asset Z if one exists. How?: Net effect of rebalancing cascade if path exists.

In the first row, M_1, you can immediately see that there is no way for an initial shock to asset A to trigger a rebalancing cascade affecting asset Z; whereas, in the rows two, three, and four there is. The middle column traces out one such path from asset A to asset Z that ends in a positive shock for asset Z in each of these three markets: M_2, M_3, M_4. But, the right column shows that there is no easy way to determine whether the net result of the entire rebalancing cascade (buy? or sell?) will be the same as in the middle column. The effect of the single path and the net effect of the entire cascade only agree in the second row, M_2.

means in plain English. Imagine the universe of all rebalancing cascades that stem from an initial positive shock to an arbitrarily small subset of assets in the market. Will every single one of these rebalancing cascades have a positive effect on asset Z after T rounds of rebalancing? Solving How means answering this question.

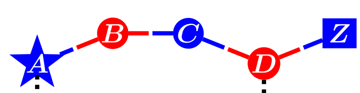
How Complexity. The proposition below gives a mathematical result that mirrors the intuition we built up in Figure 2. It says that solving the How decision problem is much harder than solving the If decision problem.

Proposition 2.3a (How Complexity)**.** How *is an* NP*-complete problem.*

Just like instances of If, the size of an instance of How is governed by the number of assets in the market, S. The complexity class NP is the set of decision problems with solutions that can be verified in polynomial time. A crossword puzzle is a good example of a problem that is hard to solve but easy to verify (Garey and Johnson, 2002). Solving this Sunday's grid might take an hour, but it will only take a second to verify your guess about the answer for 31-down once you see the answer key in next week's paper.

What does it mean for a decision problem to be NP complete? For any pair of decision problems, Prob_1 and Prob_2, we say that solving Prob_2 can be reduced to solving Prob_1 if you can solve Prob_2 by just mapping each instance of Prob_2 over to a corresponding instance of Prob_1 and then simply solving Prob_1. Intuitively, if solving Prob_2 can be reduced to solving Prob_1, then solving Prob_2 is no harder than solving Prob_1. A decision problem is NP complete if it belongs to NP and every other NP problem can be reduced to it.

Root of the Problem. Figure 3 illustrates precisely why How is so computationally intractable. Each vertical gray region denotes a separate sequence of events, starting at the top and ending at the bottom. On the left, there is a proposed path connecting asset A to asset Z that ends with a positive shock to asset Z:



But, researchers cannot immediately conclude that a positive shock to stock A will result in a rebalancing cascade that leads to a positive shock to stock Z. Knowing that this one path exists is not enough. The trouble is that assets A and D are also connected to other assets that may not belong to the original path (dotted lines). This means that the market structure could contain a secondary path.

The four gray regions to the right show how small changes in the length of this secondary path can change the rebalancing cascade's net effect on asset Z. If asset A and asset D are

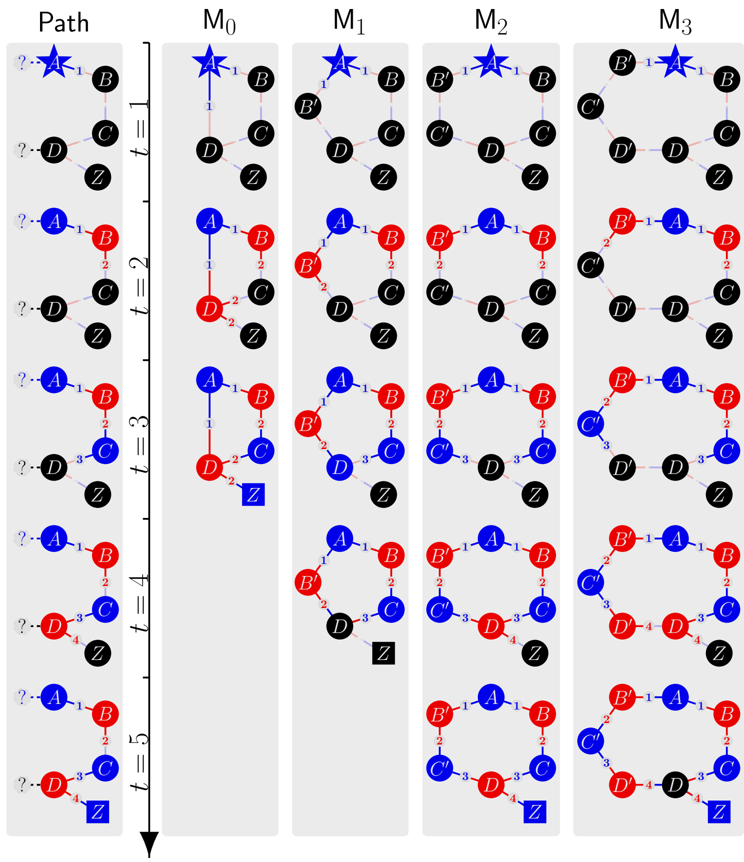


Figure 3. Root of the Problem. Each vertical gray region denotes a separate sequence of events, which starts at the top and ends at the bottom. Each node denotes an asset. Node color denotes effect of cascade: blue=positive, red=negative, black=no effect. Star: initial shock to asset A. Square: final effect for asset Z. Edges denote rebalancing rules. Blue(s)-to-red(s'): asset s' is negative neighbor to asset s. Red(s)-to-blue(s'): asset s' is positive neighbor to asset s. Path: path connecting asset A to asset Z. Location of assets A, B, C, D, and Z remain unchanged in all sequences. Dotted lines: neighbors to asset A and asset Z that could form alternate path. M_k: market structure that contains alternate path with $k \in \{0, 1, 2, 3\}$ assets separating A and D.

Left column depicts a path connecting asset A to asset Z via assets B, C, and D that results in a buy order for asset Z when ignoring the other connections that assets A and D have. Right four columns then illustrate how tiny changes in how assets A and D are connected to the rest of the network can alter the effect of a cascade starting with asset A. Cascade will result in buy orders for asset Z in markets M_0, M_2, and M_3 but not in market M_1. Moreover, you cannot tell which situation you are in just by looking at asset Z's immediate neighbors. While asset D realizes no net effect in both market M_1 and market M_3, the cascade only results in a positive net demand shock for asset Z in market M_3.

directly connected, as in M_0, then the secondary path does not matter. If there is a 1-path connecting asset A to asset D, as in M_1, then the secondary path implies that asset Z will be unaffected by the entire rebalancing cascade. But, if there is a 2-path connecting asset A to asset D, as in M_2, then the secondary path will not matter once again. If there is a 3-path connecting asset A to asset D, as in M_3, then asset Z will be positively affected by the rebalancing cascade even though asset D will be unaffected. Tiny changes in the structure of a rebalancing network can cause the resulting rebalancing cascade to affect asset Z in completely different ways.

Thus, determining how a rebalancing cascade will affect asset Z requires a detailed simulation of the entire cascade. So, finding an initial shock which negatively effects asset Z could require checking every possible ϵ-small subset. The size of this power set growth exponentially with the number of assets, S. "A running time that scales exponentially implies a harsh bound on the problems we can ever solve—even if our project deadline is as far away in the future as the Big Bang is in the past (Moore and Mertens, 2011)." For further details, see the numerical simulations in Section 2.4 (Figure 4) as well as our discussion of time complexity in the Internet Appendix.

Predicting How. Proposition 2.3a says that the problem of figuring out how every single rebalancing cascade will effect asset Z is computationally intractable. But, maybe this is an unreasonable goal. What if you only tried to figure out how rebalancing cascades will affect asset Z on average?

Problem 2.3b (MajorityHow)**.**

- *Instance: An asset* Z*, market structure* M*, time* $T = \mathrm{Poly}[S]$*, positive constant* $\epsilon > 0$*, and power set* $\hat{\mathsf{A}} \subseteq 2^{\mathsf{S}}$ *of all* ϵ*-small sets* $\mathsf{A} \subseteq \mathsf{S}$.
- *Question: Is* $\sum_{\mathsf{A} \in \hat{\mathsf{A}}} 1_{\{\mathrm{Effect}_{\mathsf{M},T}(\mathsf{A},Z)=+1\}} > |\hat{\mathsf{A}}|/2$*?*

Compared to How, the MajorityHow decision problem seems like a much closer analogue to the problem that real-world traders care about. A trader might know which funds hold each asset. They also might know the rebalancing rules involved. Given this information, they would like to determine how some asset Z will be affected by the majority of rebalancing cascades that might occur. i.e., for a given market structure, will more than half of all possible rebalancing cascades result in buy orders for asset Z?

At first, MajorityHow might seem much easier to solve than How because it does not involve finding a particular verboten instance. But, this first reaction is wrong. Proposition 2.2b shows that asset Zs with more neighbors are more likely to be hit by a rebalancing cascade. But, Proposition 2.3b shows that determining whether more than half of these rebalancing cascades will result in buy orders for asset Z is tantamount to predicting the outcome of a coin flip.

Proposition 2.3b (Predicting How)**.** MajorityHow *is an* NP*-hard problem.*

A decision problem is NP hard if every decision problem in NP can be reduced to it but the problem itself might not belong to NP. So, if MajorityHow is an NP-hard problem, then it is at least as hard as any decision problem in NP.

2.4 Numerical Simulations

We just showed that, although it is possible to predict if a rebalancing cascade starting with stock A will affect the demand for an unrelated stock Z, it is computationally infeasible to predict how this cascade will affect stock Z. As a result, we can think of the demand shocks to stock Z that come from such rebalancing cascades as noise: computational complexity makes them inherently unpredictable, and they stem from shocks that have nothing to do with stock Z's fundamentals. Since complexity theory plays a central role in this analysis, we use numerical simulations to give some intuition about the core concepts involved.

Solve Time. If can be solved in polynomial time (Proposition 2.2a) while How cannot be (Proposition 2.3a). To give a sense of what this means, in the left panel of Figure 4 we simulate markets with an increasing number of stocks using the Erdos and Rényi (1960) random-network model with connectivity parameter $\kappa = \frac{1}{2} \cdot \log S$ like in Figure 2. For each market size on the x-axis, $S \in \{100, 200, \ldots, 1{,}000\}$, we use this procedure to create 200 markets. We report the average time (dashed) as well as the longest time (solid) it takes to solve both the If problem (bottom) and the How problem (top) for each market. Since the solve time depends on the kind of computer used to do the calculation, we normalize so the average time to solve the If problem with $S = 100$ stocks represents 1 unit of time.

The left panel of Figure 4 shows that, on average, the How problem takes 1,000 times as long to solve as the If problem when there are only $S = 100$ stocks (ratio of dotted lines at left-most point). But, when there are $S = 1{,}000$ stocks, it takes roughly 500,000 times longer to solve the How problem on average (ratio of dotted lines at right-most point). In the worst case, it can take 10 billion times longer to solve the How problem in the worst case rather than the average case (ratio of solid lines at right-most point). Since the y-axis is on a logarithmic scale, these ratios will get larger and larger at an increasing rate as $S \to \infty$. We give more background on the intuitive distinction between problems that are solvable in

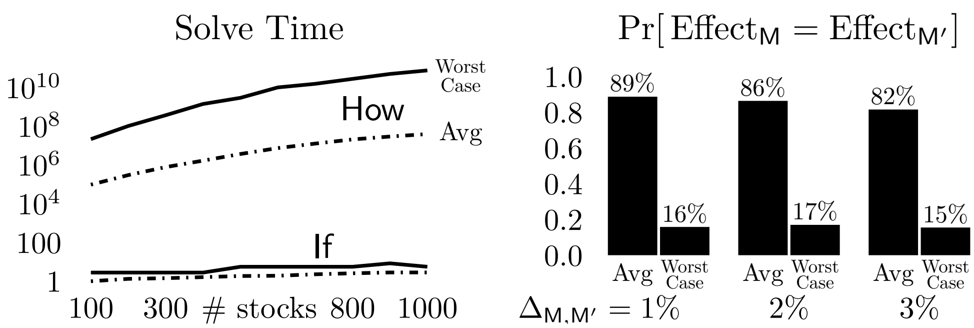


Figure 4. Assessing Complexity. Results of simulations that use the Erdos and Rényi (1960) random-network model with connectivity parameter $\kappa = \frac{1}{2} \cdot \log S$ to sample rebalancing networks. See discussion on page 12 for more details. **(Left Panel)** Top Lines, How: time needed to compute how a rebalancing cascade will affect the demand for stock Z. Bottom Lines, If: time needed to compute if a rebalancing cascade might affect the demand for stock Z. Dashed Lines, Average: Average time needed to solve each problem across 200 iterations. Solid Lines, Worst Case: Maximum time needed to solve each problem across 200 iterations. x-axis: Number of stocks used in the simulation, $S \in \{100, 200, \ldots, 1{,}000\}$. Numbers on each line are normalized so that 1 unit of time is how long it takes to solve the If problem with $S = 100$ stocks. **(Right Panel)** Probability that the effect of a rebalancing cascade on stock Z would be unchanged if a small number of edges in the rebalancing network were changed in a market with $S = 1{,}000$ stocks. $\Delta_{\mathsf{M},\mathsf{M}'} \in \{1\%, 2\%, 3\%\}$ is the fraction of edges that are different between markets M and M′. Left Bar, Average: Average fraction of stock Zs with unchanged cascade effects taken across 200 iterations. Right Bars, Worst Case: Fraction of stock Zs with unchanged cascade effects in the worst case.

polynomial time and those that are not on page 1 of the Internet Appendix.

Approximate Knowledge. Quoting a difference in solve times makes it seem like computational complexity would not be a problem if market participants had fast enough computers. But, the same mathematical considerations that make the How problem computational intractable would persist even if computers got much much faster. The How problem takes a long time to solve because tiny changes in the structure of a rebalancing network can completely change the net effect of a rebalancing cascade on stock Z.

All you have to do to solve the If problem is find one path connecting stock A to stock Z. Once you find this path, you can ignore the rest of the market structure. By contrast, there are no shortcuts to solving the How problem. You have to mechanically simulate out an entire rebalancing cascade to see what its net effect will be on stock Z. This means that it takes a prohibitively long time to solve the How problem. It also means that, if market participants get even a few links in the rebalancing network wrong, then the computational complexity of the How problem can allow these errors to blow up and completely change the net effect of a rebalancing cascade on stock Z. In other words, the computational complexity of the How problem implies that approximate knowledge of most funds' rebalancing rules will not lead to an approximately correct guess about the net demand for stock Z. If it did, then the How problem would be easy to solve.

Real-world investors only ever have approximate knowledge of most funds' rebalancing rules. Even the simplest rebalancing rules are never as simple in practice as they seem to academic economists. e.g., the S&P 500 does not just hold the largest 500 companies listed US stock exchanges. Standard & Poor's also uses a variety of other criteria to select the constituents of this index. According to Standard & Poor's methodology, "initial public offerings should be seasoned for 6 to 12 months before being considered for addition to an index." This rule lead to a bizarre situation in 2013 where Facebook had a market cap of over \$120b immediately following its IPO while still being excluded from the index.[4] These sorts of small differences in rebalancing rules can have big effects. e.g., "the \$8.6 billion iShares Edge MSCI USA Momentum ETF [MTUM] has trailed the S&P 500 Index by almost 4% since the beginning of January, putting the fund on track for its worst start to the year since it started trading in 2013…and it's all down to a quirky rule that dictates when the fund tweaks its holdings."[5] In addition, they may be hard to detect. e.g., "Vanguard Group Inc. mistakenly added shares of 11 companies, including a gun manufacturer and a private-prison operator, to its \$578 million socially responsible ETF."[6]

We can use the same simulation framework as before to illustrate this insight. The right panel of Figure 4 reports the probability that the effect of a rebalancing cascade on stock Z in market M will be unchanged if investors get a small number of edges in the rebalancing network wrong when there are $S = 1{,}000$ stocks. Let $\Delta_{\mathsf{M},\mathsf{M}'} \in \{1\%, 2\%, 3\%\}$ denote the fraction of edges that are different between markets M and M'. We simulate each market M using the Erdos and Rényi (1960) random-network model with connectivity parameter $\kappa = \frac{1}{2} \cdot \log S$ just like before.

For each simulation, we sample a market structure M and compute the effect of a rebalancing cascade starting with a shock to stock A on each stock Z in this market structure. Then, we change $\Delta_{\mathsf{M},\mathsf{M}'} \in \{1\%, 2\%, 3\%\}$ of the links in M. We next recompute the effect of the exact same cascade starting with stock A on each of the 999 stock Zs when using M' rather than M. The quantity $\Pr[\,\mathrm{Effect}_{\mathsf{M}} = \mathrm{Effect}_{\mathsf{M}'}\,]$ represents the fraction of the 999 stock

[4] CNN Business. 12/6/2013. *When Will Facebook Be Added To The S&P 500?*

[5] Bloomberg. 4/30/2019. *How Bad Timing Screwed Up Returns for a \$9 Billion Momentum Bet.*

[6] Bloomberg. 8/22/2019. *Vanguard Social-Investing Error Prompts Funds to Check Controls.*

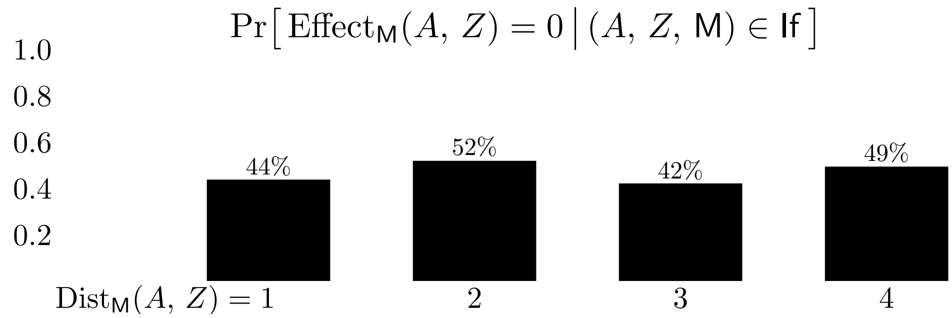


Figure 5. Effect of Distance. Results of simulations that use the Erdos and Rényi (1960) random-network model with $S = 1{,}000$ stock and connectivity parameter $\kappa = \frac{1}{2} \cdot \log S$ to sample 200 different rebalancing networks. See discussion on page 12 for more details. $\text{Dist}_{\mathsf{M}}(A, Z)$ represents the length of the shortest path from stock A to stock Z in the market structure M. If $\text{Dist}_{\mathsf{M}}(A, Z) < \infty$, then $(A, Z, \mathsf{M}) \in \mathsf{If}$. Each bar reports the probability that a rebalancing cascade starting with stock A will have a zero net effect on stock Z when $\text{Dist}_{\mathsf{M}}(A, Z) \in \{1, 2, 3, 4\}$.

Zs that have unchanged demand shocks coming from the cascade. When $\Delta_{\mathsf{M},\mathsf{M}'} = 1\%$ in a simulation, investors think that the market structure is M' when the true market structure is M, and this error only involves 1% of the links in the rebalancing network. In other words, investors in this simulation would know the true market structure with 99% accuracy. When $\Delta_{\mathsf{M},\mathsf{M}'} = 2\%$, they know the true market structure with 98% accuracy. And when $\Delta_{\mathsf{M},\mathsf{M}'} = 3\%$, they know it with 97% accuracy.

The left-most bar shows that, if investors had 99%-accurate knowledge of the rebalancing network and infinitely fast computers, then they would make correct predictions about the effect of a rebalancing cascade on the demand for 89% of stock Zs on average. But, the next bar over shows that, in the worst case, a 1% error in knowledge of the rebalancing network could lead to incorrect demand predictions for 84% of stock Zs! Thus, the right panel of Figure 4 reveals that investors with approximate knowledge of the rebalancing network are not guaranteed approximately correct answers to the How problem.

Effect of Distance. The threshold-based rebalancing rules we study have an all-or-nothing flavor. e.g., a typical fund holds the top/bottom $X\%$ of stocks when sorting on some variable like firm size, past returns, volatility, etc…This means that funds completely change their position in a stock whenever it crosses over the top/bottom $X\%$ threshold. As a result of these all-or-nothing decisions, the effect of a rebalancing cascade starting with stock A need not dissipate with distance from stock A like the waves caused by a stone thrown into water dissipate with distance from the point of impact.[7] If there is a path connecting stock A to stock Z in the rebalancing network, $(A, Z, \mathsf{M}) \in \mathsf{If}$, then a rebalancing cascade starting with a shock to stock A is just as likely to have a non-zero effect on stock Z, $\text{Effect}_{\mathsf{M}}(A, Z) \in \{-1, 1\}$, no matter how long the path is from stock A.

We show as much using numerical simulations in Figure 5. To create this figure, we simulate 200 markets containing $S = 1{,}000$ stocks using the Erdos and Rényi (1960) random-network model with connectivity parameter $\kappa = \frac{1}{2} \cdot \log S$ like before. For each market, we compute the effect of a rebalancing cascade starting with stock A on each stock Z with a connecting path to stock A. The black bars report the probability across all simulations that a cascade will have no net effect on a stock Z, $\text{Effect}_{\mathsf{M}}(A, Z) = 0$, when stock Z is $\text{Dist}_{\mathsf{M}}(A, Z)$ links away from stock A. If rebalancing cascades decayed with distance, then you would expect the height of these black bars to be rising with distance. They are not. Conditional on stock A and stock Z being connected at all, there is no statistical difference between $\Pr[\,\text{Effect}_{\mathsf{M}}(A, Z) = 0\,]$ when $\text{Dist}_{\mathsf{M}}(A, Z) = 1$ (immediate neighbors) relative to when $\text{Dist}_{\mathsf{M}}(A, Z) = 4$ (unrelated).

We study ETF rebalancing decisions in our empirical analysis. And, in the real world, not every ETF rebalancing decision looks like the ones we analyze in our theoretical framework. Unlike what we show in Figure 5, rebalancing cascades involving such alternative rules will generally have effects that decay with distance from stock A for reasons we discuss in the following subsection. This is fine for our analysis. While ETFs sometimes use other kinds of rebalancing rules, they also often rebalance using the one-for-one threshold-based rules in our model. Figure 5 shows that this particular kind of rebalancing rule leads to cascades that do not decay with distance from stock A.

2.5 Key Ingredients

We have just seen that predicting how a rebalancing cascade will affect an asset's demand (buy? or sell?) with accuracy better than a coin flip is an NP-hard problem, even if you are fully rational and funds are following simple deterministic rebalancing rules. As a result, market participants will treat the demand shocks to stock Z that come from long rebalancing cascades as noise. These demand shocks are fundamentally unpredictable and they have nothing to do with stock Z's fundamentals since they start with an initial shock to an entirely unrelated stock A. We now describe three key features that make rebalancing cascades so hard to predict. The fact that these three features are commonly observed

[7] We would like to thank an anonymous referee for suggesting this evocative metaphor.

throughout modern financial markets suggests that rebalancing cascades are an important noise-generating mechanism.

Alternation. First, rebalancing cascades are only hard to predict if they involve alternating sequences of buy and sell orders. In our model, a positive shock to stock A causes a negative shock to a neighboring stock B. Whereas, a negative shock to stock A would cause a positive shock to a neighboring stock B. This is what we mean by an alternating sequence of buy and sell orders. In a world where a positive shock to asset A can only result in a positive shock to asset B or a negative shock to asset A can only result in a negative shock to asset B, there would be no alternation and predicting how asset Z would be affected would be easy. All you would have to do is check if the cascade would affect Z at all. Without alternation, there is only one way to be affected by a cascade. So, if Z is affected, you know exactly how. Thus, without alternation, the How problem is equivalent to solving the 'If?' problem.

Proposition 2.5a (Necessity of Alternation)**.** *Without alternation,* How *is solvable in polynomial time.*

There are cascade-like phenomena without alternation—i.e., phenomena in which the cascade carries uni-directional shocks. For example, think about bank runs. During a bank run, depositors are choosing whether to withdraw their money—the decision is about whether to sell or do nothing at all. There is no alternation involved. As a result, equilibrium demand in these models behaves in a predictable way depending on whether some critical value has been reached (Diamond and Dybvig, 1983).

However, alternation is an essential part of what it means to rebalance a portfolio. When a fund rebalances, it necessarily exchanges an existing position in one asset for a new position in another. The asset being added receives a positive shock while the asset being dropped receives a negative shock. Thus, our results show how a seemingly innocuous fact about each individual rebalancing decision—that it delivers a positive shock to one asset and a negative shock to another—can make a big difference when it comes to predicting the outcome of long rebalancing cascades.

Interacting Paths. Second, rebalancing cascades are only hard to predict in a market structure that involves cancellation due to interacting paths. It is important that different cascade paths have the potential to cancel each other out, as highlighted in Figures 1 and 3. We say that a cancellation has occurred at time $(t + 1)$ for stock s if $x_{s,t} = \pm 1$ and $x_{s,t+1} = 0$ when updating the cascade using Equations (4a) and (4b). Thus, making cancellations off limits means restricting the space of allowable market structures, M. If M does not contain any cancellations following an initial shock to any ϵ-small collection of assets $\hat{\mathsf{A}}$, then there can be at most one path connecting any pair of assets in M. Market structures that do not permit cancellations will look stringy and contain no loops.

To see why the absence of interacting paths would matter, think about what would happen if every asset in the market had no more than two neighbors. In this setting, there is no way for a single asset to be affected by a rebalancing cascade more than once. So, if there existed a path connecting asset A to asset Z, then you could easily determine how a rebalancing cascade starting with asset A would affect the demand for asset Z by just counting the number of assets in the path from A to Z. If that number was odd, the shock to asset Z would be positive (Equation 8); if even, it would be negative (Equation 9).

Proposition 2.5b (Necessity of Interacting Paths)**.** *Without interacting paths,* How *is solvable in polynomial time.*

Again, we feel that interaction between various funds' rebalancing rules is a natural part of modern financial markets. There is no central-planning committee that limits the number of funds holding a single asset. There is nothing stopping 20 different smart-beta ETFs from holding the same company at the same time for different reasons.[8] Thus, the associated collection of rebalancing rules will contain market structures with feedback loops. It is these loopy instances where different paths interact with one another that make solving How computationally intractable. They make it so that a researcher cannot predict the outcome of a rebalancing cascade starting with stock A on stock Z by checking a single path from stock A to stock Z. As illustrated in Figure 3, the interaction between this path and other paths from stock A to stock Z can result in cancellations that completely change the net effect on stock Z.

Thresholds. Third, rebalancing cascades are only hard to predict if funds follow threshold-based trading rules. For example, it is important that the PowerShares S&P 500 Low-Volatility ETF [SPLV] tracks a benchmark consisting of only the 100 lowest-volatility stocks on the S&P 500. An alternative approach would be to track a low-volatility benchmark that included at least a few shares of all stocks but relatively more shares of the lowest-volatility stocks. In the first case, an arbitrarily small change in the volatility of stock s can move it from 101st to 100th place on the low-volatility leaderboard and force SPLV to exit its entire position in stock s. In the second, a small change in the volatility of stock s

[8] SeekingAlpha. 6/27/2017. *Smart Beta ETFs Love These Stocks.*

would only ever lead to a small change in SPLV's position in stock s.

The indicator functions $1_{\{|\mathsf{In}^+_{s,t+1}|>|\mathsf{In}^-_{s,t+1}|\}}$ and $1_{\{|\mathsf{In}^+_{s,t+1}|<|\mathsf{In}^-_{s,t+1}|\}}$ in Equation (4a) capture the notion of thresholds. They imply that there is no difference between the strength of the various incoming links in the rebalancing network. i.e., it does not matter how much the volatility of the 100th least volatile stock increased. As long as its volatility increased enough to push the stock out of the 100 least volatile stocks, the SPLV fund will have to dump its position in the stock and build a new position in what was previously the 101st stock on the low-volatility leaderboard. All that matters is crossing the threshold. How much a stock's fundamentals change does not matter. Going from 100th to 150th on the low-volatility leaderboard would have had the same effect as going from 100th to 101st.

If low-volatility funds hold only the 100 stocks with the lowest volatility (a threshold-based rule), then stock s is either 100th or 101st on the low-volatility leaderboard. But, if low-volatility funds do not use threshold-based rules and just hold slightly more shares of lower volatility stocks, then a flip-flop of stock s and stock s' in the rankings from 100th to 101st would not result in much rebalancing if these stocks' respective volatilities did not change that much. Thus, without threshold-based rebalancing rules, we could rewrite Equations (4a) and (4b) as a single linear system:

$$\underset{[S\times 1]}{\Delta \boldsymbol{x}_{t+1}} = (\underset{[S\times S]}{\mathsf{In}^-} - \underset{[S\times S]}{\mathsf{In}^+})\,\underset{[S\times 1]}{\Delta \boldsymbol{x}_t} \tag{12}$$

Without thresholds, a small shock to stock s will result in a small opposing shock to its neighbors; a large shock to stock s will result in a large opposing shock to its neighbors. As a result, the effect of a rebalancing cascade in which funds do not use thresholds could be any real number rather than $x_{s,t} \in \{-1, 0, 1\}$ as before.

Let $\text{SmoothEffect}_{\mathsf{M},T}(\mathsf{A}, Z)$ denote the effect on stock Z's demand of the rebalancing cascade defined on page 9 when Step 3 is computed using the new smooth updating rule in Equation (12) rather than the two rules in Equations (4a) and (4b). Below we define a new How-like decision problem associated with this sort of smoothly rebalancing cascade.

Problem 2.5c (SmoothHow)**.**

- *Instance: An asset Z, market structure M, time $T = \text{Poly}[S]$, positive constant $\epsilon > 0$, and power set $\hat{\mathsf{A}} \subseteq 2^{\mathsf{S}}$ of all ϵ-small sets $\mathsf{A} \subseteq \mathsf{S}$.*
- *Question: Is there some $\mathsf{A} \in \hat{\mathsf{A}}$ such that $\text{SmoothEffect}_{\mathsf{M},T}(\mathsf{A}, Z) < 0$?*

"When there are interacting paths and rebalancing cascades involve alternating sequences of buy and sell orders, different branches of a rebalancing cascade stemming from the same initial shock can interfere with one another. If there is no way for a small change in fundamentals to get amplified, such as when it causes a stock to cross a threshold, then the effects of long rebalancing paths from stock A to stock Z will not be that large for stock Z. Thus, without threshold-based rebalancing rules, longer cascade paths in the cascade can effectively be ignored for the same reason that AR(1) impulse-response functions get exponentially weaker at longer horizons. The ability to ignore such longer paths makes SmoothHow much easier to solve than How."

Proposition 2.5c (Necessity of Thresholds)**.** SmoothHow *is solvable in polynomial time.*

Compare and contrast this result with Proposition 2.3b, which says that, in the presence of threshold-based rebalancing rules, it is impossible to predict whether the answer is '*Yes.*' to even slightly more than half of all How instances. We know that many funds use threshold-based rebalancing rules. A typical stat-arb trading strategy will have the form, '*Buy the top* 30% *and sell the bottom* 30% *of stocks when sorting on X*', where X is some variable that predicts the cross-section of expected returns. This means that funds will completely change their position in a stock whenever it crosses over the top/bottom 30% threshold. Our goal is not to explain why funds use these trading rules; instead, our goal is to point out that noise is a natural consequence of the fact that they do.

3 Empirical Evidence

The noise-generating mechanism we propose in this paper predicts how noise should vary across assets. Specifically, stocks on the cusp of more rebalancing thresholds should experience more noise because they are more susceptible to getting hit by a rebalancing cascade. In this section, we study rebalancing cascades in the exchange-traded fund (ETF) market to provide empirical support for this prediction. Throughout our analysis, we use `teletype font` to indicate variables used in regressions or reported in summary statistics.

3.1 Exchange-Traded Funds

Our theoretical model characterizes rebalancing cascades following an initial publicly observed shock to some asset A. So, to bring the model to the data, we need to choose a group of funds to study as well as a collection of initial shocks. We now address the first of these choices. We use data on the end-of-day positions of U.S. ETFs from January 1st, 2011 to December 31st, 2017. Our data comes from ETF Global, and we only consider ETFs in this database that rebalance more than once a quarter.

Market Selection. There are three reasons why we chose to study the ETF market. First, it closely resembles the environment that we model in Section 2. There are now more ETFs than stocks.[9] The sheer number and variety of these so-called *smart-beta* ETFs divides popular opinion.[10] To be sure, niche ETFs tend to be smaller than broad value-weighted market ETFs, such as the well-known SPDR S&P 500 ETF. But, there are a lot of these niche funds now. As Steve Sachs, head of capital markets for ETFs at Goldman, describes: "Our institutional clients and our advisor clients were coming to us and asking us, 'Can you deliver these strategies in an ETF?"'[11] Moreover, even the rebalancing activity of niche ETFs can affect a stock's underlying characteristics because ETFs execute the bulk of their trades during the final 20-to-30 minutes prior to close.[12]

Second, the model applies to a setting where fund managers follow simple deterministic rebalancing rules. ETF managers do not deviate from their stated benchmarks like mutual-fund or hedge-fund managers do (Madhavan, 2016; Ben-David et al., 2017). ETF trading volume primarily comes from rebalancing activity just prior to market close. The numbers are stark: "37% of New York Stock Exchange trading volume now happens in the last 30 minutes of the session, according to JPMorgan. The chief culprit is the swelling exchange-traded fund industry. . . ETFs are essentially investment algorithms of varying degrees of complexity, and typically automatically rebalance their holdings at the end of the day."[13] This end-of-day trading is how ETFs make sure that there is very little difference between the market value of their end-of-day holdings and the value of their stated benchmark.

Third, the model relies on the fact that the trading activity of one fund can trigger other funds' trading rules. To observe these distortions empirically, we need to be able to see a fund's holdings on a daily basis. We can observe end-of-day portfolio positions for ETFs. Note that other papers in the ETF literature, such as Ben-David et al. (2017), impute each ETF's daily portfolio position from its end-of-quarter financial statements. This is a perfectly reasonable approach for answering some research questions, but it will not work in our setting. We are interested in how the rebalancing decisions of different index funds interact with one another over the course of a few days. We cannot study these interactions by imputing a fund's daily holdings from end-of-quarter reports.

Our theoretical analysis shows that, even when funds are using simple deterministic

[9] Bloomberg. 5/16/2017. *Mutual Funds Ate the Stock Market. Now ETFs Are Doing It.*
[10] New York Times. 06/22/2017. *An Index-Fund Evangelist Is Straying From His Gospel.*
[11] CNBC. 4/21/2019. *Goldman Sachs ETF chief breaks down new investment trend.*
[12] Wall Street Journal. 03/14/2018. *What's the Biggest Trade on the NYSE? The Last One.*
[13] Financial Times. 3/17/2017. *Machines and Markets: 5 Areas To Watch.*

rebalancing rules, the resulting cascades can still be formally intractable to predict. Our goal in the empirical analysis is to document evidence of this noise-generating mechanism in real-world financial markets. And, for this goal, it is fine that not all ETF rebalancing decisions look like the one-for-one threshold-based rules that we model in our theoretical framework. What is important for our purposes is that many ETF issuers do track their target portfolio in the simple way we model.

Variable Construction. For each ETF, $f \in \{1, \ldots, F\}$, ETF Global provides data on assets under management, $\texttt{AUM}_{f,t}$, and relative portfolio weights in stock s, $\Omega_{f,s,t}$, at the end of each trading day t from January 1st, 2011 to December 31st, 2017. 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 fth ETF holds on day t is $Q_{f,s,t} = \frac{1}{P_{s,t}} \times (\Omega_{f,s,t} \cdot \texttt{AUM}_{f,t})$. Total ETF trading volume for stock s on day t is given by $\texttt{ETFvlm}_{s,t} = \sum_{f=1}^{F} | Q_{f,s,t} - Q_{f,s,t-1} |$. We restrict our sample to only include ETFs that rebalance more than once a quarter. So, when viewing our results, you should have in mind the PowerShares S&P 500 Low-Volatility ETF rather than the SPDR S&P 500 ETF. We also exclude leveraged ETFs from our sample. We do this to emphasize that we are studying how rebalancing cascades can transmit an initial shock to stock A to an unrelated stock Z in an unpredictable way and not how leveraged ETFs predictably amplify initial shocks to stock A (Ivanov and Lenkey, 2018). We do not otherwise restrict, winsorize, or filter the data.

We use this end-of-day ETF-holdings data to create two main variables of interest. The first is ETF rebalancing volume. This requires a little bit of subtlety because, while most ETF trading volume each day is due to rebalancing decisions, some ETF trading volume is not. Here is how we isolate ETF rebalancing volume. We start with a simple accounting identity: the total amount of money that an ETF has invested in a stock must be equal to the price of the stock times the number of shares the ETF holds, $P_{s,t} \cdot Q_{f,s,t} = \Omega_{f,s,t} \cdot \texttt{AUM}_{f,t}$. Rearranging this accounting identify yields an expression for the number of shares that the ETF holds on a given day:

$$Q_{f,s,t} = \tfrac{1}{P_{s,t}} \times (\Omega_{f,s,t} \cdot \texttt{AUM}_{f,t}) \tag{13a}$$

$$= \tfrac{1}{P_{s,t}} \times (\Omega_{f,s,t-1} + \Delta\Omega^{rebal}_{f,s,t} + \Delta\Omega^{vw}_{f,s,t}) \cdot \texttt{AUM}_{f,t} \tag{13b}$$

In the second line, we have broken the ETF's portfolio weight on stock s into three components, $\Omega_{f,s,t} = \Omega_{f,s,t-1} + \Delta\Omega^{rebal}_{f,s,t} + \Delta\Omega^{vw}_{f,s,t}$.

The first component, $\Omega_{f,s,t-1}$, is the ETF's portfolio weight on the previous day. The second component, $\Delta\Omega^{rebal}_{f,s,t}$, is the change in the ETF's portfolio weight due to rebalancing decisions—e.g., due to the stock getting added to or deleted from the ETF's benchmark index. The third component, $\Delta\Omega^{vw}_{f,s,t}$, is the change due to value weighting. If $R_{s,t}$ denotes the return of stock s and $R_{bmk,t}$ denotes the return on the ETF's value-weighted benchmark on day t, then we can express this third component as follows using observable data (see the internet appendix for details):

$$\Delta\Omega^{vw}_{f,s,t} = (R_{s,t}/R_{bmk,t} - 1) \cdot \Omega_{f,s,t-1} \tag{14}$$

If the change in the price of a stock leads to a change in a value-weighted index, then there is no need for an ETF tracking this index to rebalance. The weight of the stock in the basket owned by the ETF will be automatically adjusted. So, this third term will be zero if a fund is tracking a value-weighted index. But, there are a lot of ETFs that do not track value-weighted indexes.[14] So, we still need to include this term.

We are only interested in changes in an ETF holdings due to rebalancing decisions. Creations and redemptions will cause the size of the fund to change, $\texttt{AUM}_{f,t} \neq \texttt{AUM}_{f,t-1}$. But, any resulting transactions will get executed as in-kind transfers for tax reasons, which would mean that $\Delta\Omega^{rebal}_{f,s,t} = 0$. Likewise, if a stock appreciates in value, then it will receive a higher weight in an ETF's portfolio, but the ETF will not actually have to rebalance its position in the stock. So, we would have a situation where $\Delta\Omega^{rebal}_{f,s,t} = 0$.

So, to compute the amount of a stock's trading volume coming specifically from ETF rebalancing decision, we first calculate each ETF's predicted holdings for the stock on day t, $\bar{Q}_{f,s,t} = \frac{1}{P_{s,t}} \times (\Omega_{f,s,t-1} + \Delta\Omega^{vw}_{f,s,t}) \cdot \texttt{AUM}_{f,t}$. Then, for each stock, we sum up the difference between every ETF's actual end-of-day holdings and this prediction:

$$\texttt{ETFrebal}_{s,t} = \textstyle\sum_{f=1}^{F} \left| Q_{f,s,t} - \bar{Q}_{f,s,t} \right| \tag{15}$$

We use this as our daily measure of ETF rebalancing volume. Market participants clearly talk about the effects of daily ETF rebalancing volume. "For ETFs, rebalancing can trigger matching in-then-out trades that resemble the peaks and valleys on an electrocardiogram, earning them the nickname 'heartbeat flows'. The trades are most often seen in funds that track specialized indexes where companies are frequently added or dropped from the portfolio."[15] The end-of-day burst in trading volume coming from ETF rebalancing is not just something that happens at quarter end. It is an everyday phenomenon.

To measure the direction of ETF order flow (buy? or sell?), we also compute a measure

[14] Wall Street Journal. 9/8/2019. *Your Diversified ETF Might Now Be Anything But.*

[15] The Wall Street Journal. 6/5/2019. *ETF 'Heartbeats' Show Influence of Indexes.*

of the order imbalance in each stock's ETF rebalancing volume:

$$\texttt{ETFimbal}_{s,t} = \begin{cases} \sum_{f=1}^{F} \frac{Q_{f,s,t} - \bar{Q}_{f,s,t}}{\texttt{ETFrebal}_{s,t}} & \text{if } \texttt{ETFrebal}_{s,t} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

This variable lies on the interval $[-1, 1]$. If $\texttt{ETFimbal}_{s,t} = -1$, every share of stock s traded by ETFs for rebalancing reasons resulted in a sell order. Whereas, if $\texttt{ETFimbal}_{s,t} = +1$, every share of ETF rebalancing volume resulted in a buy order. Table 1 provides summary statistics describing the ETF market and the stocks they hold.

3.2 M&A Announcements

Having explained why we chose to study the ETF market, we now describe why we use M&A announcements for our set of initial shocks. We refer to the target of an M&A announcement as stock A. Our data on M&A deals come from Thomson Reuters SDC Mergers and Acquisitions database. We use all deals that involve publicly traded target firms with an announcement date between January 1st, 2011 and December 31st, 2017. There are 1119 such deals in our seven-year sample period, yielding an average of 14.3 announcements per month. We report additional summary statistics in the internet Appendix.

Effect on Rebalancing. M&A announcements are a natural choice for our initial shocks for several reasons. First, there is solid empirical evidence that the target of an M&A announcement realizes a sharp price change immediately following the announcement (Andrade et al., 2001). Second, many ETFs have clauses in their charters which state that they cannot hold clear takeover targets (Madhavan, 2016). The combination of these two forces often results in other changes in target-stock fundamentals as well, such as increases in volatility. While acquirers do not choose their M&A targets at random, the exact day that a deal is announced—Wednesday vs. Thursday—can be taken as exogenous.

Consistent with these findings, Table 2 shows that ETFs rebalance their position in stock A on the day that stock A gets revealed as the target of an M&A deal. Let t_A denote the day stock A is announced as an M&A target. We create a panel dataset containing the ETF rebalancing volume for each stock A during the 26-day window $t \in \{t_A - 20, \ldots, t_A + 5\}$. Then, we regress stock A's ETF rebalancing volume on indicator variables for the number of days until the M&A announcement:

$$\texttt{ETFrebal}_{A,t} = \alpha_A + \alpha_{mmyy} + \beta \cdot 1_{\{t=t_A-1\}} + \gamma \cdot 1_{\{t=t_A\}} + \delta \cdot 1_{\{t=t_A+1\}} + \varepsilon_{A,t} \tag{17}$$

Above, α_A and α_{mmyy} denote stock-A and month-year fixed effects respectively.

The first column in Table 2 shows that ETF rebalancing volume for stock A rises by

156.59% on the exact day that stock A is announced as an M&A target. The second column then shows results for the same regression specification after including controls for stock A's lagged total trading volume. The point estimate for the effect of an M&A announcement on ETF rebalancing volume in stock A hardly changes when moving from the first to the second column. The third column shows the results of a similar specification that also includes additional indicator variables for the days $(t_A - 2)$, $(t_A - 3)$, $(t_A - 4)$, and $(t_A - 5)$ prior to stock A's announcement. This column reveals that there is no pre-trend in ETF managers' reaction to the M&A announcement. This timing is consistent with the fact that ETF managers do not have any discretion when it comes to deviating from their benchmark index overnight.

Placebo Test. The fifth and final column of Table 2 shows the results of the same regression specification using placebo announcement dates for each stock A. We randomly re-assign the announcement date t_A to some other point in our sample period prior to each stock A's actual announcement. Consistent with the idea that it is the M&A announcement itself that is causing the jump in ETF rebalancing volume, we find that there is no jump in ETF rebalancing volume on these placebo dates.

3.3 Experimental Design

The model predicts that, following an initial shock to stock A, i) stock Zs with more neighbors in the ETF rebalancing network should be involved in ETF rebalancing cascades more often, but ii) it should not be possible to predict the direction of these cascades' effect on a stock Z's demand. Here is how we test these two predictions.

Unrelated Stock Zs. To start with, we need to make sure that any ETF rebalancing activity we measure is due to ETF rebalancing cascades and not some omitted variable affecting both stock A and stock Z. So, we create a separate panel for each M&A announcement containing the set of stock Zs that are unrelated to stock A during the 26-day window $t \in \{t_A - 20, \ldots, t_A - 1, t_A, t_A + 1, \ldots, t_A + 5\}$ around stock A's announcement.

For a stock Z to be unrelated to stock A, the target of an M&A announcement, these two stocks have to be twice removed in the network of ETF rebalancing decisions. Stock Z cannot have been rebalanced at any point during the last month by any ETF that also held stock A during the last month. Moreover, if stock B and stock A were both held at any point during the last month by the same ETF, then stock Z cannot have been rebalanced during the last month by any ETF that also held stock B during that time period. In other words, the chain of ETF rebalancing decisions from stock A to stock Z has to be $A \to B \to C \to Z$ or longer. Because there are smart-beta ETFs tracking things like large-cap, value, and industry, this criteria implies that stock A and stock Z do not have any similar factor exposures and do not share any well-known firm characteristics.

We then combine these separate datasets—one for each of the 1119 M&A announcements in our sample—into a single panel dataset indexed by M&A announcement, stock Z, and date. Because the same stock Z can be affected by ETF rebalancing cascades starting with different initial stock As, we index the rows of this dataset with the subscript $Z,t|A$.

Diff-in-Diff Approach. We study this panel dataset using a diff-in-diff approach. The model looks at rebalancing cascades following an initial shock to stock A. So, the first difference will capture whether or not this initial shock to stock A—i.e., the announcement that stock A is the target of an M&A deal—has occurred yet. We define $\texttt{afterAncmt}_{t|A}$ as an indicator variable for the five days immediately after the announcement about stock A:

$$\texttt{afterAncmt}_{t|A} = \begin{cases} 1 & \text{if } t \in \{t_A + 1, \ldots, t_A + 5\} \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

We write $\texttt{afterAncmt}_{t|A}$ rather than $\texttt{afterAncmt}_{Z,t|A}$ because the post-announcement period is the same for all stock Zs that are unrelated to the target stock A.

The model then makes predictions about the differential effect of a rebalancing cascade on stock Zs depending on their number of neighbors in the ETF rebalancing network. The second difference captures whether a particular stock Z has lots of neighbors in the ETF rebalancing network. We say that stock s is a neighbor to stock Z if an ETF that currently holds stock s also rebalanced its position in stock Z during the previous month.

For each M&A announcement, we use this definition to split the set of stock Zs into two subsets: those on the cusp of an above-median number of ETF rebalancing thresholds (i.e., stock Zs with lots of neighbors in the ETF rebalancing network) and those on the cusp of a below-median number of ETF rebalancing thresholds. Let $\texttt{manyNhbr}_{Z|A}$ be an indicator variable for whether or not stock Z has an above-median number of neighbors relative to all other stocks which are unrelated to stock A:

$$\texttt{manyNhbr}_{Z|A} = \begin{cases} 1 & \text{if stock } Z \text{ has an above-median number of neighbors} \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

We calculate the number of neighbors for each stock Z using data from the month prior to each M&A announcement. So, this indicator variable does not vary during window surrounding each announcement, which is why we write $\texttt{manyNhbr}_{Z|A}$ rather than $\texttt{manyNhbr}_{Z,t|A}$. But, because we calculate the median number of stock-Z neighbors separately for each M&A

announcement, this indicator variable can vary across M&A announcements for the same stock Z. The exact same stock Z can have an above-median number of neighbors in the ETF rebalancing network relative to one M&A announcement but a below-median number relative to another. This is why we write $\mathtt{manyNhbr}_{Z|A}$ rather than $\mathtt{manyNhbr}_{Z}$.

Proposition 2.2b predicts that stock Zs with more neighbors in the ETF rebalancing network will be more likely to be hit by an ETF rebalancing cascade. We use the following regression specification to test this prediction:

$$\begin{aligned}\mathtt{ETFrebal}_{Z,t|A} = \alpha_{\#\mathrm{nhbr}|A} + \alpha_Z + \beta \cdot \mathtt{afterAncmt}_{Z|A} \\ + \gamma \cdot \{\mathtt{afterAncmt}_{t|A} \times \mathtt{manyNhbr}_{Z|A}\} + \varepsilon_{Z,t|A}\end{aligned} \tag{20}$$

Here is what each of the resulting coefficients means. First, consider the two fixed effects: $\alpha_{\#\mathrm{nhbr}|A}$ and α_Z. We include announcement×`#nhbr` fixed effects because the same initial announcement about stock A might result in either a large or a small ETF rebalancing cascade depending on subtleties of how the ETF rebalancing network is wired up. The fact that we are using announcement×`#nhbr` fixed effects rather than just announcement fixed effects means that we are including separate coefficients for the average level of rebalancing volume among stock Zs with one neighbor in the days around stock A's announcement, the average level of rebalancing volume among stock Zs with two neighbors in the days around stock A's announcement, the average level of rebalancing volume among stock Zs with three neighbors in the days around stock A's announcement, and so on... In addition to the announcement×`#nhbr` fixed effects, we include stock-Z fixed effects to account for the fact that ETFs might always trade some stocks than others.

Next, let's consider the coefficients on the first difference, β. This coefficient captures the rise in ETF rebalancing volume for all stock Zs in the five days immediately after an M&A announcement about an unrelated stock A. Because an ETF rebalancing cascade has the potential to affect the demand for all stocks—i.e., even stock Zs with few neighbors—we should expect the average ETF rebalancing volume of all stock Zs to rise in five days after the M&A announcement. In other words, we should expect to estimate $\beta > 0$ in Equation (20). After all, there are more benchmark indexes than stocks in modern financial markets. Most stocks in the market will be connected by at least one direct neighbor (Erdos and Rényi, 1960), and it is possible for these stocks Zs to be affected by an ETF rebalancing cascade. The model predicts it is more likely for stock Zs with many neighbors to be hit.

The coefficient γ captures how much more the ETF rebalancing volume increases for stock Zs with many neighbors than for stock Zs with few neighbors in the five days

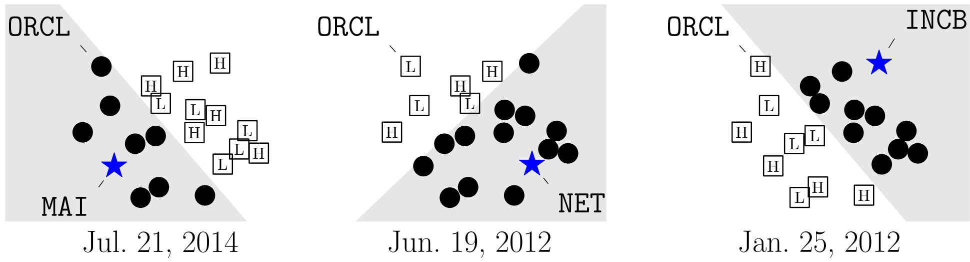


Figure 6. Time-Varying Comparison Groups. Each panel depicts the the same set of stocks during 3 different M&A announcements: Owens & Minor's purchase of Medical Action Industries [MAI] announced on Jul. 21, 2014; Sonus Networks' purchase of Network Equip Technologies [NET] announced on Jun. 19, 2012; and, Old National Bancorp's purchase of Indiana Community Bancorp [INCB] announced on Jan. 25, 2012. The target of each M&A announcement, stock A, is denoted by a blue star. Each black circle denotes a stock that is related to stock A at the time of the announcement. Each white square denotes a stock that is unrelated to stock A at the time of the announcement. This is the set of stock Zs. Unrelated stocks that are neighbors with an above-median number of other stocks are labeled with an "H"; whereas, those that are neighbors with a below-median number are labeled with an "L". Oracle Corp. is a related stock in the left panel, a below-median stock Z in the middle panel, and an above-median stock Z in the right panel.

immediately after stock A's M&A announcement. The first key prediction of the model is that we should estimate $\gamma > 0$ in Equation (20). By contrast, Proposition 2.3b suggests that it should not be possible to predict the direction (buy? or sell?) of the resulting demand shocks. So, the second key prediction of the model is that, if we replace ETF rebalancing volume with ETF order imbalance in Equation (20),

$$\begin{aligned}\mathtt{ETFimbal}_{Z,t|A} = \alpha_{\#\mathrm{nhbr}|A} + \alpha_Z + \beta \cdot \mathtt{afterAncmt}_{Z|A} \\ + \gamma \cdot \{\mathtt{afterAncmt}_{t|A} \times \mathtt{manyNhbr}_{Z|A}\} + \varepsilon_{Z,t|A},\end{aligned} \tag{21}$$

then we should estimate $\gamma = 0$. We do not include the level effect for $\mathtt{manyNhbr}_{Z|A}$ in the specification because it gets subsumed by the announcement×`#nhbr` fixed effects.

Source of Identification. At this point, it is important to pause and spell out the source of our identification in these regressions. We want to emphasize that we are not making an assumption that stock Zs with many neighbors are similar stock Zs with few neighbors in the ETF rebalancing network. In fact, there are good reasons to expect these stocks to be different. M&A announcements are just one kind of initial shock that might trigger an ETF rebalancing cascade. So, if you really believe that ETF rebalancing cascades generate noise, then you should expect stock Zs with an above-median number of neighbors to always have more ETF rebalancing volume than below-median stock Zs since these stocks will be hit by more ETF rebalancing cascades. Summary statistics for above- and below-median stock Zs can be found in Table A2 in the internet appendix.

Instead, our identification is coming from the timing of the M&A announcements. We are starting out with an initial M&A announcement about stock A and then looking at the set of stock Zs that are totally unrelated to stock A. We are going to show that, even though these stock Zs are totally unrelated to stock A, i) ETF rebalancing volume immediately after stock A's announcement increases more for above-median stock Zs since these stocks are more likely to be hit by a rebalancing cascade, and ii) this increase in ETF rebalancing volume is just as likely to consist of buy orders as of sell orders. We recognize that above-median stock Zs are different than below-median stock Zs on average, but there is no reason to expect the size of this difference to increase immediately after an M&A announcement about an unrelated stock A in the absence of a cascade.

This identification strategy raises two kinds of concerns. First, you might be worried that something else about the set of unrelated stock Zs is changing at the time of the M&A announcements. This is something we can test in the data. Figure A2 in the internet appendix shows that, although above- and below-median stock Zs tend to have different amounts of ETF trading activity, this difference is constant in the run up to each M&A announcement. The figure also shows that the same statement holds for other stock Z characteristics. Again, we should expect to observe differences between above- and below-median stock Zs since above-median stock Zs are always more likely to be hit by ETF rebalancing cascades. It is the fact that these differences suddenly change in the wake of an unrelated M&A announcement that is important.

The other concern you might have is about an omitted variables problem. Perhaps something else is happening at the same time as each M&A announcement, and it is this omitted variable that is causing ETFs to trade above-median stock Zs differently. This omitted-variables problem would be a major concern if our data contained a small number of M&A announcements and the number of neighbors for each stock Z remained constant over time. If this were this case, then there might plausibly be some alternative story for why ETFs happened to trade a particular group of stock Zs differently at a few particular moments in time. But, this is not at all what our data looks like. We have lots of M&A announcements. The exact same stock can be an above-median stock Z relative to one M&A announcement while simultaneously being a below-median stock Z relative to another as shown in Figure 6. Thus, our findings cannot be explained by ETFs always trading some stock Zs differently than others. Any omitted variable would have to account for why ETFs suddenly change their rebalancing behavior for only the stock Zs that have an above-median number of neighbors relative to a particular stock A in the five days immediately after that stock A's M&A announcement.

Why Not Track Each Step? Finally, among economists, it is taken almost as an article of faith that empirical analysis is best run using the most micro-level data possible. So, you might be surprised that we have not just traced out the precise buy-sell-buy-sell sequence involved in each ETF rebalancing cascade. However, there is a good reason why we have not done this: it would fundamentally ignore a central insight of our theoretical analysis—namely, that it is computationally infeasible to make predictions about how a rebalancing cascade will affect each stock's demand. As illustrated via numerical simulations in Figure 4, overlooking even a few links in the ETF rebalancing network can reverse a cascade's effect on the demand for most stock Zs. While our data on the end-of-day holdings of each ETF is good, it is not perfect. No data is. It would be a miracle if our data were not missing at least a few links.

If we take this insight to heart, then it is clear that we need to run our empirical analysis using well-chosen macro-level variables rather than the most micro-level data possible. Even if it is not practical to track the precise buy-sell-buy-sell sequence of ETF rebalancing decisions, it is relatively easy to measure how many ETF rebalancing thresholds each stock Z is on the cusp of. By analogy, even if it is not possible to keep track of the location and momentum of every single gas molecule in a 1m^3 box, it is relatively easy to measure macro-level variables like the pressure and temperature inside the container. We focus our empirical analysis on a particular kind of fund—exchange-traded funds (ETFs)—following a particular kind of initial shock—M&A announcements—because this setting provides a nice laboratory where we can measure these well-chosen macro-level variables.

3.4 Estimation Results

We now provide empirical evidence that i) while an unrelated stock Z with many neighbors is more likely to be hit by a rebalancing cascade, ii) the resulting demand shock is just as likely to be composed of buy orders as of sell orders.

ETF Rebalancing Volume. Table 3 describes how cascades affect ETF rebalancing volume by reporting the estimated coefficients for the regression specification in Equation (20). The first column shows that ETF rebalancing volume for all stock Zs tends to rise by $\beta = 2.64\%$ on average in the five days immediately after an M&A announcement about an unrelated stock A. But, the third column shows that this growth is concentrated among stock Zs that have many neighboring stocks in the network defined by ETF rebalancing rules. Consistent with the model, we find that ETF rebalancing volume is $\gamma = 2.06\%$ higher for above-median stock Zs than for below-median stock Zs in the five days immediately following an M&A announcement about an unrelated stock A. The second and fourth columns of Table 3 confirm that the sudden spike in the ETF rebalancing volume is not due to a general run-up in trading volume. When we include stock Z's total trading volume in our regression specification, our point estimate for γ remains largely unchanged.

We run two different kinds of placebo tests to make sure that our estimate of $\gamma \approx 2\%$ is due to ETF rebalancing cascades and not some omitted variable. The results of the first placebo test can be found in the fifth column of Table 3. For this column, we re-estimated the regression specification in Equation (20) using data during the 26-day window surrounding $t_A - 30$ rather than t_A. If we shift stock A's announcement date forward by 30 days, then we can be certain that there was no initial M&A announcement and thus no subsequent ETF rebalancing cascade. When we do this, our point estimate for γ shrinks by a factor of ten, from 2.06% to 0.20%, and becomes statistically indistinguishable from zero. What's more, this lack of statistical significance is being completely driven by the smaller coefficient estimate. The lack of statistical significance is not due to a lack of power. Our standard error on γ is roughly the same, 0.35 vs. 0.32, in both the true and the placebo samples. We describe the second placebo test in Table A3 found in the internet appendix.

ETF Order Imbalance. Next, in Table 4, we give evidence that the extra ETF rebalancing volume experienced by above-median stock Zs following an unrelated M&A announcement is no more likely to be made up of buy orders than of sell orders. This table reports the estimated coefficients for the regression specification in Equation (21). The point estimate of $\gamma = 0.74\%$ with a standard error of 0.51% in second column reveals that there is no statistically measurable difference between the ETF order imbalance of above- and below-median stock Zs following an M&A announcement. Taken together, this evidence suggests that, while it is possible to predict which stock Zs are likely to be affected by a ETF rebalancing cascade, it is much harder to predict how these stock Zs will be affected by the resulting demand shock.

Price Pressure. Although ETF rebalancing demand is no more likely to be positive than negative on average, Table 5 reveals that it still has a significant effect on prices. Columns (1) and (2) show the results of replacing the dependent variable in Equation (20) with the daily returns of each stock Z. On average, the prices of above-median stock Zs look just like the prices of below-median stock Zs following an initial shock to an unrelated stock A.

Columns (3) and (4) of Table 5 then show the differential effect of very positive or very negative demand from ETF rebalancing on the prices of these same subgroups. In Table 3 we show that above-median stock Zs tend to realize larger demand shocks.[16] As a result, the prices of above-median stock Zs rise by 14bps per day more than those of below-median stock Zs in response to order flow composed of a larger fraction of buy orders. And the prices of these stocks shrink by an extra 21bps per day in response to order flow with a higher proportion of sell orders.

If the demand coming from ETF rebalancing cascades is noise, then we would expect that the resulting price pressure should be transient. In column (5), this is exactly what we find. This column reports results for the exact same regression specification as in columns (1)-(4) only now using a slightly longer window following each M&A announcement. Instead of only using the 5 days following each announcement, column (5) uses data on the 10 days following each announcement. When we do this, the price pressure coming from ETF rebalancing cascades on stock Z gets cut in half. On the positive side, the coefficient drops from 14bps to 7bps. On the negative side, the coefficient shrinks from −21bps to −9bps. This is exactly the sort of decay in the price-impact coefficient that we would expect if the demand from ETF rebalancing cascades were noise.

Economic Magnitude. We have just seen evidence that ETF rebalancing cascades generate unpredictable demand shocks by focusing on the effects of ETF rebalancing cascades in the wake of M&A announcements. Thus, since not all ETF rebalancing cascades start with an initial M&A announcement, these results represent a lower bound for the total amount of demand noise generated by ETF rebalancing cascades. We use a collection of panel regressions to get a better sense of the total amount of noise the could be produced by ETF rebalancing cascades irrespective of the initial shock.

We regress each stock's log ETF rebalancing volume normalized by its standard deviation over the past twelve months, $\hat{\sigma}_{s,t}$, on the number of neighbors this stock has in the ETF rebalancing network in thousands:

$$\texttt{ETFrebal}_{s,t}/\hat{\sigma}_{s,t} = \alpha + \beta \cdot \texttt{\#nbrs}_{s,t} + \varepsilon_{s,t} \tag{22}$$

The first column of Table 6 reports that $\beta = 0.40$, which implies that a 1σ increase in the

[16] It is possible that the demand shocks coming from different paths in a rebalancing cascade cancel out, but such exact cancellations are unlikely to begin with and become increasingly unlikely as the number of paths increases. Thus, they are not a concern for above-median stock Zs. By analogy, it is possible for $N \geq 4$ coin flips to contain exactly $N/2$ heads; however, this is not the most likely outcome, $\Pr[\#heads \neq N/2] > \Pr[\#heads = N/2]$. And such an exact outcome becomes less likely as more coins are flipped, $N \to \infty$.

number of neighbors a stock has is associated with a 0.4σ increase in a stock's log ETF rebalancing volume. Moreover, the second column of Table 6 shows that this point estimate remains statistically significant even if we include fixed effects for the number of ETFs holding the stock, suggesting the effect is operating through a stock's position in the network and not merely through the number of ETFs that directly hold the stock.

Since not all ETF rebalancing volume is demand noise, this estimate represents an upper bound for the amount of noise coming from ETF rebalancing cascades. However, not all rebalancing cascades involve ETFs. So, the results in Table 6 do not represent a hard cap on the amount of demand noise generated by many funds rebalancing. Instead, they show that, when operating in the realm of ETFs, rebalancing cascades have the potential to generate an economically meaningful amount of demand noise.

4 Conclusion

> *"To generate randomness, we humans flip coins, roll dice, shuffle cards, or spin a roulette wheel. All these operations follow direct physical laws, yet casinos are in no risk of losing money. The complex interaction of a roulette ball with the wheel makes it computationally impossible to predict the outcome of any one spin, and each result is indistinguishable from random."* —Fortnow (2017)

This paper proposes an analogous explanation for where noise comes from in financial markets. An asset's demand might appear random, not because individual investors are actually behaving randomly, but because it is too computationally complex to predict how a wide variety of simple deterministic trading rules will interact with one another. We show theoretically how computational complexity can generate noise by modeling a particular kind of trading-rule interaction: rebalancing cascades. Then, we give empirical evidence on how this noise-generating mechanism predicts which assets will have the most noise using data on the end-of-day holdings of ETFs.

This insight offers news ways to test existing models. Consider your favorite limits-to-arbitrage model. In the past, if you wanted to test this model, then you would just look for situations where arbitrageurs were most constrained. But, in order for the limits of arbitrage to bind, there must also be a non-fundamental shock (Chinco, 2018). Our noise-generating mechanism says that stocks with more neighbors in the ETF rebalancing network will realize more of these shocks. So, you can now check whether the implications of the limits-to-arbitrage model are strongest for constrained stocks with the most noise.

Technical Appendix

Definition (Binary String)**.** Let $\{0,1\}^\star = \cup_{n=0,1,\ldots}\{0,1\}^n$ be the set of binary strings.

Definition (Problem Solving)**.** Let $\mathsf{Prob} \in \{0,1\}^\star$ denote a decision problem. An algorithm $\mathrm{F} : \{0,1\}^\star \mapsto \{0,1\}$ solves Prob (a.k.a., decides membership in Prob) if for every instance $i \in \{0,1\}^\star$ we have that

$$i \in \mathsf{Prob} \quad \Leftrightarrow \quad \mathrm{F}(i) = 1$$

Problem A (stCon)**.**

- *Instance: A directed graph* G *and two vertices* (s,t).
- *Question: Is there a path from* s *to* t*?*

Theorem A (Wigderson, 1992)**.** stCon *is solvable in polynomial time.*

Definition (Reduction)**.** Let Prob_1 and Prob_2 denote two decision problems. We say that Prob_2 is (Karp, 1972) reducible to Prob_1 if there exists a polynomial-time algorithm $\mathrm{F} : \{0,1\}^\star \mapsto \{0,1\}^\star$ such that

$$i \in \mathsf{Prob}_2 \quad \Leftrightarrow \quad \mathrm{F}(i) \in \mathsf{Prob}_1$$

Proof (Proposition 2.2a)**.** If $\hat{\mathsf{S}}$ contains a single asset, then If and stCon are the same problem—there is a trivial reduction from If to stCon. Both involve finding a path from one node in a directed network to another. What's more, each K-path to asset Z is evaluated separately. For example, in the market described by Figure 1, the path described in Equation (8) exists with or without the path described by Equation (9). This means that if $(Z, \mathsf{M}, T, \{s\}) \in \mathsf{If}$ and $(Z, \mathsf{M}, T, \{s'\}) \in \mathsf{If}$, then $(Z, \mathsf{M}, T, \{s, s'\}) \in \mathsf{If}$. Thus, we do not need to check every single subset $\hat{\mathsf{S}} \subseteq \mathsf{S}$ separately. To see which subsets of assets are connected to asset Z, we can just check which assets are connected to asset Z. This is reducible to solving $(S-1)$ separate instances of stCon, which is doable in polynomial time because stCon itself if solvable in polynomial time (Wigderson, 1992). □

Proof (Equation 10)**.** Suppose M contains S assets and was generated using connectivity parameter $\kappa > 0$. If $(s, s') \in \mathsf{S}^2$, then asset s' will be a positive neighbor to asset s with probability κ/S. Because the outcome is determined independently for each asset $s' \in \mathsf{S}$, the probability that asset s has exactly n positive neighbors is

$$\Pr(N_s^+ = n \,|\, S) = \binom{S}{n} \cdot (\kappa/S)^n \cdot (1 - \kappa/S)^{S-n}$$

This is the probability of n successes in S independent Bernoulli trials, which implies

$$N_s^+ \sim \text{Binomial}(\kappa/S, S)$$

So, given the additional restriction that $\kappa = \mathrm{O}[\log S]$, we know that as $S \to \infty$

$$N_s^+ \sim \text{Poisson}(\kappa, S)$$

since the Binomial distribution converges to Poisson as $S \to \infty$ for small κ. □

Proof (Proposition 2.2b)**.** Let $C_s \in \{\texttt{True}, \texttt{False}\}$ be an indicator variable for whether or not an asset s is connected to the giant component of the random graph induced by M. We can write

$$\begin{aligned}\Pr[(Z, \mathsf{M}, T, \{s\}) \in \mathsf{If} \,|\, N_Z = n] &= \Pr[(C_s = \texttt{True}) \wedge (C_Z = \texttt{True}) \,|\, N_Z = n] \\ &= \Pr[C_s = \texttt{True}] \cdot \Pr[C_Z = \texttt{True} \,|\, N_Z = n]\end{aligned}$$

The second line implies that $\mathrm{E}[\hat{\mathrm{S}}_{\max}(Z, \mathsf{M}, T)]$ will be increasing in N_Z if and only if $\mathrm{E}[C_Z|N_Z = n]$ is increasing in n since the path connecting each asset $s \in \mathsf{S}$ to asset Z can be evaluated independently. Bayes' rule implies

$$\mathrm{E}[C_Z|N_Z = n] = \left(\frac{\Pr[N_Z = n|C_Z = \texttt{True}]}{\Pr[N_Z = n]}\right) \times \mathrm{E}[C_Z]$$

$\Pr[N_Z = n|C_Z = \texttt{True}]/\Pr[N_Z = n]$ is increasing in n. So, we can conclude that $\mathrm{E}[C|N = n]$ is increasing in n. □

Definition (Complexity Class NP)**.** Let Prob denote a decision problem, and let $|i|$ denote the size of instance i. We say that $\mathsf{Prob} \in \mathsf{NP}$ if there exists a polynomial-time Turing machine M such that

$$i \in \mathsf{Prob} \quad \Leftrightarrow \quad \exists\ w \in \{0, 1\}^{\mathrm{Poly}(|i|)} \text{ s.t. } \mathrm{M}(i, w) = 1$$

The string w is known as the 'witness' or 'proof' that $i \in \mathsf{Prob}$.

Definition (Hardness)**.** Let CC denote a complexity class, such as NP. Prob is hard with respect to CC if every decision problem in CC can be reduced to Prob.

Definition (Completeness)**.** Let CC denote a complexity class. We say that Prob is complete with respect to CC if both i) $\mathsf{Prob} \in \mathsf{CC}$ and ii) Prob is CC hard.

Problem B (3Sat)**.**

- *Instance: A Boolean formula defined over N input variables*

$$\mathrm{F} : \{\texttt{True}, \texttt{False}\}^N \mapsto \{0, 1\}$$

  where some clauses contain 3 *variables.*
- *Question: Is there an assignment* $\mathbf{x} \in \{\texttt{True}, \texttt{False}\}^N$ *such that* $\mathrm{F}(\mathbf{x}) = 1$*?*

Theorem B (Cook, 1971)**.** 3Sat *is an* NP*-complete problem.*

Corollary. If Prob is reducible to 3Sat, then Prob is NP complete.

Proof (Proposition 2.3a)**.** We show that How is NP complete by reducing it to 3Sat. There are two steps to the proof.

Step 1: First, create variables to track of the state of the rebalancing cascade:

- For each possible value of $(x_{s,t}, \Delta x_{s,t})$,

$$k \in \{(0{,}0), (1{,}1), (1{,}0), (0, -1), (-1, -1), (-1{,}0), (0{,}1)\}$$

  define $\alpha(k)_{s,t} = 1_{\{(x_{s,t}, \Delta x_{s,t}) = k\}}$ for each asset $s \in \mathsf{S}$.

- For each pair of asset $(s, s') \in \mathsf{S}^2$ such that $s \neq s'$ define

$$\beta^+_{s,s',t} = 1_{\{s' \in \mathsf{Out}^+_{s,t}\}} \qquad \text{and} \qquad \beta^-_{s,s',t} = 1_{\{s' \in \mathsf{Out}^-_{s,t}\}}$$

- For each pair of assets $(s, s') \in \mathsf{S}^2$ such that $s \neq s'$ define

$$\gamma^+_{s',s,t+1} = 1_{\{s \in \mathsf{In}^+_{s',t+1}\}} \qquad \text{and} \qquad \gamma^-_{s',s,t+1} = 1_{\{s \in \mathsf{In}^-_{s',t+1}\}}$$

- For each asset $s \in \mathsf{S}$ define

$$\delta^+_{s,t+1} = 1_{\{u_{s,t+1}=1\}} \qquad \text{and} \qquad \delta^-_{s,t+1} = 1_{\{u_{s,t+1}=-1\}}$$

Total number of new variables is polynomial in S.

Step 2: Encode constraints on variables in conjunctive-normal form clauses. There are two kinds of constraints to consider.

- First, there are constraints that impose variable consistency. e.g., we cannot have both $\alpha(0{,}0)_{s,t} = 1$ and $\alpha(1{,}1)_{s,t} = 1$ at the same time, $(\overline{\alpha(0{,}0)}_{s,t} \vee \overline{\alpha(1{,}1)}_{s,t})$.
- Second, there are constraints that encode the rebalancing cascade updating rules. e.g., if stock s has one negative neighbor, s', and one positive neighbor, s'', then the rebalancing-cascade rules are encoded in four different clauses:

δ^+_s	$\lambda^+_{s',s}$	$\lambda^-_{s'',s}$	Violated Clause
0	0	0	✓
0	0	1	✓
0	1	0	⊗ $(\delta^+_s \vee \bar{\lambda}^+_{s',s} \vee \lambda^-_{s'',s})$
0	1	1	✓
1	0	0	⊗ $(\bar{\delta}^+_s \vee \lambda^+_{s',s} \vee \lambda^-_{s'',s})$
1	0	1	⊗ $(\bar{\delta}^+_s \vee \lambda^+_{s',s} \vee \bar{\lambda}^-_{s'',s})$
1	1	0	✓
1	1	1	⊗ $(\bar{\delta}^+_s \vee \bar{\lambda}^+_{s',s} \vee \bar{\lambda}^-_{s'',s})$

Again, the total number of new clauses is polynomial in S.

Whenever asset s has both positive and negative neighbors, some of these clauses involve 3 variables. Thus, we have a polynomial reduction of How to $\mathsf{3Sat}$. □

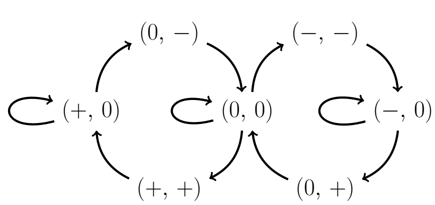


Figure 7. State Diagram. All possible ways that a single asset could move between the 7 possible values of $(x_{s,t}, \Delta x_{s,t})$ in successive rounds of a rebalancing cascade. Arrows denote transitions. Loops denote unchanged values in successive rounds.

Definition (Complexity Class PP)**.** Let Prob denote a decision problem, and let $r \in \{0, 1\}^\star$ denote an arbitrarily long sequence of random bits. We say that $\mathsf{Prob} \in \mathsf{PP}$ if there exists a polynomial-time randomized algorithm F such that

$$i \in \mathsf{Prob} \quad \Leftrightarrow \quad \Pr_r[\, \mathrm{F}(i, r) = 1 \mid i \notin \mathsf{Prob}\,] > 1/2$$

Problem C (Majority)**.**

- *Instance: A Boolean formula defined over N input variables*

$$\mathrm{F} : \{\texttt{True}, \texttt{False}\}^N \mapsto \{0, 1\}$$

- *Question: Is* $\sum_{\mathbf{x} \in \{\texttt{True},\texttt{False}\}^N} \mathrm{F}(\mathbf{x}) > 2^{N-1}$*?*

Theorem C (Gill, 1977)**.** NP $\subseteq$ PP*, and* Majority *is a* PP*-complete problem.*

Corollary. Let Prob denote any decision problem. If Prob is reducible to Majority, then Prob is PP hard.

Proof (Proposition 2.3b)**.** The proof of Proposition 2.3a showed how to reduce instances of How into Boolean formulas. So, since Majority is defined in terms of Boolean functions, the same reduction converts instances of MajorityHow into instances of Majority. Hence, because MajorityHow is a PP-complete problem (Gill, 1977), the corollary above implies that MajorityHow is an NP-hard problem. □

Problem D (2Sat)**.**

- *Instance: A Boolean formula defined over N input variables*

$$\mathrm{F} : \{\texttt{True}, \texttt{False}\}^N \mapsto \{0, 1\}$$

where no clause contains more than 2 *variables.*
- *Question: Is there an assignment* $\mathbf{x} \in \{\texttt{True}, \texttt{False}\}^N$ *such that* $\mathrm{F}(\mathbf{x}) = 1$*?*

Theorem D (Cook, 1971)**.** 2Sat *is solvable in polynomial time.*

Proof (Proposition 2.5a)**.** If there is no alternation, then assets only have positive neighbors. So, an asset Z will be affected by an initial shock to the assets in A if and only if there is a path from asset $s \in \mathsf{A}$ connecting to asset Z. Without alternation, there is no way for two different paths in a rebalancing cascade to interfere with one another. Since within a single path, each asset has only 0 (asset A) or 1 (all other assets) incoming links at any point in time, there would be no need to create clauses with more than two variables in the proof of Proposition 2.3a. Thus, without alternation, How is reducible to 2Sat. This reduction implies it is solvable in polynomial time (Cook, 1971). □

Proof (Proposition 2.5b)**.** If there is no interaction between paths, then there is either a single path from any asset s to asset Z or no such path. After all, if there is more than one path, then these two paths would define a closed loop. As a result, no asset can have more than 1 incoming link. And so, the rebalancing cascade rules can be encoded using clauses with no more than 2 variables as in the proof of Proposition 2.5a. Thus, without interacting paths, How is reducible to 2Sat. This reduction implies that it is solvable in polynomial time (Cook, 1971). □

Proof (Proposition 2.5c)**.** Computing $\mathrm{SmoothEffect}_{\mathsf{M},T}(\mathsf{A}, Z)$ for all (A, Z) pairs involves computing the impulse response of a positive shock to a linear system (cf. Equation 12), $(\mathsf{In}^- - \mathsf{In}^+)^T \Delta \boldsymbol{x}_0$, which involves only matrix multiplication. A polynomial number of matrix multiplications can be computed in polynomial time (Natarajan, 1995). □

Summary Statistics

Panel A. ETF Level

	Avg	Sd	25%	50%	75%
$\mathtt{count}_t$	1073.3	168.2	914	1017	1249
$\mathtt{\#benchmark}_t$	896.9	141.7	769	836	1045
$\mathtt{AUM}_t$	1390.0	651.8	1060.9	1485.4	1701.0
$\mathtt{\#stock}_{f,t}$	247.0	499.2	30	78	254
$\mathtt{AUM}_{f,t}$	859.6	5381.3	4.8	30.8	217.6

Panel B. Stock Level

	Avg	Sd	25%	50%	75%
$\mathtt{\#ETFholding}_{s,t}$	476.1	453.7	88	366	713
$\mathtt{\#nhbr}_{s,t}$	3.4	1.1	3.3	3.8	4.0
$\mathtt{ETFvlm}_{s,t}$	13.2	3.4	11.7	13.4	14.7
$\mathtt{ETFrebal}_{s,t}$	13.1	3.5	11.5	13.2	14.6
$\mathtt{ETFimbal}_{s,t}$	−1.2	35.6	−19.9	0.0	19.4
$\mathtt{return}_{s,t}$	1.1	11.5	−4.2	1.1	6.4
$\mathtt{mcap}_{s,t}$	5.9	23.0	0.2	0.8	3.2
$\mathtt{vlm}_{s,t}$	15.5	2.0	14.3	15.7	16.9
$\mathtt{amihud}_{s,t}$	45.6	253.8	0.1	0.2	2.8
$\mathtt{spread}_{s,t}$	46.3	89.3	4.2	10.2	38.1

Table 1. Panel A. Fund-level summary statistics. $\mathtt{count}_t$: number of ETFs in sample. $\mathtt{\#benchmark}_t$: number of benchmarks used by these ETFs. $\mathtt{AUM}_t$: total assets under management for all ETFs in \$bil. $\mathtt{\#stock}_{f,t}$: number of stocks held by an ETF. $\mathtt{AUM}_{f,t}$: assets under management in \$mil. **Panel B.** Stock-level summary statistics. $\mathtt{\#ETFholding}_{s,t}$: number of ETFs that hold a stock. $\mathtt{\#nhbr}_{s,t}$: number of neighboring stocks in the ETF rebalancing network in 1000s. $\mathtt{ETFvlm}_{s,t}$: ETF trading volume each month on a log scale. $\mathtt{ETFrebal}_{s,t}$: ETF rebalancing volume each month on a log scale. $\mathtt{ETFimbal}_{s,t}$: signed ETF rebalancing volume divided by total ETF rebalancing volume reported in %. **Panel B.** Characteristics of each stock. $\mathtt{return}_{s,t}$: return in % per month. $\mathtt{mcap}_{s,t}$: market capitalization in \$bil. $\mathtt{vlm}_{s,t}$: total trading volume each month on a log scale. $\mathtt{amihud}_{s,t}$: Amihud (2002) illiquidity measure computed daily within a given month in bp per \$1mil order. $\mathtt{spread}_{s,t}$: average bid-ask spread as a fraction of midpoint during month in bps.

ETF Rebalancing Volume, Stock A

Dependent Variable: $\texttt{ETFrebal}_{A,t}$

	Actual Announcements				Placebo
	(1)	(2)	(3)	(4)	(5)
$1_{\{t=t_A+1\}}$	94.35⋆⋆⋆ (7.07)	86.92⋆⋆⋆ (7.65)	97.28⋆⋆⋆ (7.26)	88.98⋆⋆⋆ (7.81)	−14.13 (9.09)
$1_{\{t=t_A\}}$	156.59⋆⋆⋆ (6.58)	153.33⋆⋆⋆ (6.64)	159.59⋆⋆⋆ (6.80)	156.28⋆⋆⋆ (7.00)	11.21 (9.52)
$1_{\{t=t_A-1\}}$	27.64⋆⋆⋆ (7.63)	24.00⋆⋆⋆ (7.76)	30.82⋆⋆⋆ (7.91)	27.19⋆⋆⋆ (8.19)	17.18⋆ (9.71)
$1_{\{t=t_A-2\}}$			32.39⋆⋆⋆ (8.12)	30.86⋆⋆⋆ (8.33)	−7.82 (14.66)
$1_{\{t=t_A-3\}}$			14.27⋆ (7.97)	11.39 (8.09)	−3.53 (10.60)
$1_{\{t=t_A-4\}}$			14.78⋆ (7.84)	11.86 (8.22)	−2.45 (10.12)
$1_{\{t=t_A-5\}}$			1.22 (7.08)	−1.43 (7.32)	−4.19 (12.36)
$\texttt{vlm}_{A,t-1}$		1.74 (1.45)		2.04 (1.48)	5.54 (3.38)
$\texttt{vlm}_{A,t-2}$		−2.42 (1.49)		−1.96 (1.52)	1.53 (3.02)
$\texttt{vlm}_{A,t-3}$		−5.93⋆⋆⋆ (1.24)		−5.75⋆⋆⋆ (1.25)	−4.48⋆⋆ (2.66)
Month-Year FE	Y	Y	Y	Y	Y
Stock-A FE	Y	Y	Y	Y	Y
R^2	78.7%	78.8%	78.8%	78.8%	80.3%
Observations	15,459	15,434	15,459	15,434	15,434

Table 2. Effect of initial M&A announcement about stock A on ETF rebalancing volume for the same stock A. Sample: January 1st, 2011 to December 31st, 2017. Each column represents results from a separate regression using daily data on the 26-day window surrounding each M&A announcement, $\{t_A - 20, \ldots, t_A - 1, t_A, t_A + 1, \ldots, t_A + 5\}$. Columns (1)-(4) report results for actual announcements. Column (5) reports results using a randomly assigned announcement date for each M&A target in our sample. $\texttt{ETFrebal}_{A,t}$: dependent variable is the ETF rebalancing volume for stock A on date t reported on a base-e logarithmic scale. Coefficient estimate of 1 indicates a 1% increase in a stock's daily ETF rebalancing volume. $1_{\{t=t_A-h\}}$: indicator variable that is one if an observation was made h days prior to stock A's announcement date. $\texttt{vlm}_{A,t}$: total trading volume for stock A on day t reported on a base-e logarithmic scale. Numbers in parentheses are standard errors clustered by stock. Statistical significance: ⋆ = 10%, ⋆⋆ = 5%, and ⋆⋆⋆ = 1%. *Reads: "Stock A has* 156.59% *more ETF rebalancing volume on the day it is announced as an M&A target."*

ETF Rebalancing Volume, Stock Z

Dependent Variable: $\texttt{ETFrebal}_{Z,t|A}$

	Actual Announcements				Placebo
	(1)	(2)	(3)	(4)	(5)
$\texttt{afterAncmt}_{t\|A} \times \texttt{manyNhbr}_{Z\|A}$			2.06⋆⋆⋆ (0.35)	1.99⋆⋆⋆ (0.35)	0.20 (0.32)
$\texttt{afterAncmt}_{t\|A}$	2.64⋆⋆⋆ (0.15)	2.30⋆⋆⋆ (0.15)	1.63⋆⋆⋆ (0.24)	1.32⋆⋆⋆ (0.24)	1.57⋆⋆⋆ (0.25)
$\texttt{vlm}_{Z,t\|A}$		24.13⋆⋆⋆ (0.50)		24.13⋆⋆⋆ (0.50)	24.99⋆⋆⋆ (0.50)
Announcement×#nhbr FE	Y	Y	Y	Y	Y
Stock-Z FE	Y	Y	Y	Y	Y
R^2	76.2%	76.4%	76.2%	76.4%	75.9%
Observations		23,256,554			23,574,154

Table 3. Effect of an initial M&A announcement about stock A on the ETF rebalancing volume for an unrelated stock Z. Sample: January 1st, 2011 to December 31st, 2017. Each column presents the results of a separate regression using daily data on the 26-day window surrounding each M&A announcement, $\{t_A - 20, \ldots, t_A - 1, t_A, t_A + 1, \ldots, t_A + 5\}$. Columns (1)-(4) report results for actual announcements. Column (5) reports results using a placebo sample where each M&A announcement is re-assigned to date $t_A - 30$. $\texttt{ETFrebal}_{Z,t|A}$: dependent variable is ETF rebalancing volume for stock Z on date t following M&A announcement about stock A on a base-e logarithmic scale; coefficient of +1 indicates a 1% per day increase in a stock's ETF rebalancing volume. $\texttt{afterAncmt}_{t|A}$: indicator variable that is one during the five days following an M&A announcement about stock A. $\texttt{manyNhbr}_{Z|A}$: indicator variable that is one if stock Z has an above-median number of neighbors in the ETF rebalancing network relative to the M&A announcement about stock A. $\texttt{vlm}_{Z,t|A}$: total trading volume for stock Z on a given day reported on a base-e logarithmic scale. Numbers in parentheses are standard errors clustered by stock Z. Statistical significance: ⋆ = 10%, ⋆⋆ = 5%, and ⋆⋆⋆ = 1%. *Reads: "In the five days after an M&A announcement about stock A, unrelated stock Zs with an above-median number of neighbors in the ETF rebalancing network realize* 2.06% *per day more ETF rebalancing volume than unrelated stock Zs with a below-median number of neighbors."*

ETF Order Imbalance, Stock Z

Dependent Variable: $\mathtt{ETFimbal}_{Z,t|A}$

	(1)	(2)
$\mathtt{afterAncmt}_{t\|A} \times \mathtt{manyNhbr}_{Z\|A}$	0.73 (0.51)	0.74 (0.51)
$\mathtt{afterAncmt}_{t\|A}$	0.00 (0.78)	0.02 (0.78)
$\mathtt{vlm}_{Z,t\|A}$		$-1.58^{\star\star\star}$ (0.21)
Announcement×#nhbr FE	Y	Y
Stock-Z FE	Y	Y
R^2	2.4%	2.4%
Observations	23,264,687	

Table 4. Effect of initial M&A announcement about stock A on the ETF order imbalance for an unrelated stock Z. Sample: January 1st, 2011 to December 31st, 2017. Each column presents the results of a separate regression using daily data on the 26-day window surrounding each M&A announcement, $\{t_A - 20, \ldots, t_A - 1, t_A, t_A + 1, \ldots, t_A + 5\}$. $\mathtt{ETFimbal}_{Z,t|A}$: dependent variable is ETF order imbalance for stock Z on date t following M&A announcement about stock A; coefficient estimate of +1 indicates a 1% per day increase in a stock's ETF order imbalance. $\mathtt{afterAncmt}_{t|A}$: indicator variable that is one during the five days following an M&A announcement about stock A. $\mathtt{manyNhbr}_{Z|A}$: indicator variable that is one if stock Z has an above-median number of neighbors in the ETF rebalancing network relative to the M&A announcement about stock A. $\mathtt{vlm}_{Z,t|A}$: total trading volume for stock Z on a given day reported on a base-e logarithmic scale. Numbers in parentheses are standard errors clustered by stock Z. Statistical significance: $\star$ = 10%, $\star\star$ = 5%, and $\star\star\star$ = 1%. *Reads: "Although ETF rebalancing volume is higher for stocks Zs with many neighbors than for stock Zs with few neighbors in the five days after an M&A announcement about stock A, this ETF rebalancing volume is no more likely to consist of buy orders than of sell orders."*

Price Pressure, Stock Z

Dependent Variable: $\texttt{return}_{Z,t|A}$

	[−20, + 5] (1)	[−20, + 5] (2)	[−20, + 5] (3)	[−20, + 5] (4)	[−20, +10] (5)
$\texttt{posETFimbal}_{Z,t\|A} \times \texttt{manyNhbr}_{Z\|A} \times \texttt{afterAncmt}_{t\|A}$			14.40⋆⋆⋆ (1.22)	14.54⋆⋆⋆ (1.23)	7.28⋆⋆⋆ (0.85)
$\texttt{manyNhbr}_{Z\|A} \times \texttt{afterAncmt}_{t\|A}$	−0.48 (0.33)	−0.41 (0.33)	1.36⋆⋆⋆ (0.42)	1.29⋆⋆⋆ (0.42)	0.80⋆⋆ (0.36)
$\texttt{negETFimbal}_{Z,t\|A} \times \texttt{manyNhbr}_{Z\|A} \times \texttt{afterAncmt}_{t\|A}$			−22.35⋆⋆⋆ (1.29)	−22.12⋆⋆⋆ (1.29)	−9.20⋆⋆⋆ (0.91)
$\texttt{posETFimbal}_{Z,t\|A} \times \texttt{manyNhbr}_{Z\|A}$			0.49 (0.47)	0.88⋆ (0.47)	0.74⋆ (0.44)
$\texttt{negETFimbal}_{Z,t\|A} \times \texttt{manyNhbr}_{Z\|A}$			0.04 (0.46)	0.23 (0.46)	−0.92⋆⋆ (0.45)
$\texttt{vlm}_{Z,t\|A}$		11.69⋆⋆⋆ (0.37)		11.62⋆⋆⋆ (0.37)	11.37⋆⋆⋆ (0.37)
afterAncmt × ETFimbal group FE	Y	Y	Y	Y	Y
Announcement×#nhbr FE	Y	Y	Y	Y	Y
Stock-Z FE	Y	Y	Y	Y	Y
R^2	2.3%	2.5%	2.6%	2.7%	2.3%
Observations	23,256,554	23,256,554	23,256,554	23,256,554	28,064,917

Table 5. Effect of ETF order imbalance on stock Z's returns in the days after an M&A announcement about an unrelated stock A. Sample: January 1st, 2011 to December 31st, 2017. Each column presents results of a separate regression using daily data on a narrow window surrounding each M&A announcement. Columns (1)-(4) use data during the 26-day window $[t_A - 20, t_A + 5]$. Column (5) uses data during the 31-day window $[t_A - 20, t_A + 10]$. $\texttt{return}_{Z,t|A}$: dependent variable is the return of stock Z on date t; coefficient of +1 indicates a 1bps increase in a stock's daily return. $\texttt{afterAncmt}_{t|A}$: indicator variable that is one during days following an M&A announcement about stock A. $\texttt{manyNhbr}_{Z|A}$: indicator variable that is one if stock Z has an above-median number of neighbors in the ETF rebalancing network relative to the M&A announcement about stock A. $\texttt{posETFimbal}_{Z,t|A}$: indicator variable that is one if stock Z has an above-75%tile ETF order imbalance on day t. $\texttt{negETFimbal}_{Z,t|A}$: indicator variable that is one if stock Z has a below-25%tile ETF order imbalance on day t. $\texttt{vlm}_{Z,t|A}$: total trading volume for stock Z each day reported on a base-e logarithmic scale. Numbers in parentheses are standard errors clustered by stock Z. Statistical significance: ⋆ = 10%, ⋆⋆ = 5%, and ⋆⋆⋆ = 1%. *Reads: "While ETF demand does not affect a stock's price on average, buy orders due to ETF rebalancing decisions have a strong positive effect and sell orders due to ETF rebalancing decisions have a strong negative effect."*

Monthly Panel Regressions, All Stocks

Dependent Variable:	$\texttt{ETFrebal}_{s,t}$		$\texttt{amihud}_{s,t}$		$\texttt{spread}_{s,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)
$\texttt{\#nhbr}_{s,t}$	$0.40^{\star\star\star}$ (0.01)	$0.12^{\star\star\star}$ (0.01)	$-0.15^{\star\star\star}$ (0.01)	$-0.03^{\star\star\star}$ (0.01)	$-0.12^{\star\star\star}$ (0.01)	$-0.07^{\star\star\star}$ (0.01)
$\texttt{\#ETFholding}$ FE	*N*	*Y*	*N*	*Y*	*N*	*Y*
Month-Year FE	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*
Stock FE	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*
R^2	85.5%	87.7%	79.0%	80.1%	54.7%	55.0%
Observations	309,853		311,033		311,023	

Table 6. Relationship between the number of neighbors that a stock has in the ETF rebalancing network and its liquidity. Sample: January 2012 to December 2017. Each column represents results from a separate regression using stock×month observations. $\texttt{ETFrebal}_{s,t}$: log ETF rebalancing volume for stock s in month t divided by its standard deviation in the previous 12 months; coefficient of +1 represents a 1sd increase in ETF rebalancing volume. $\texttt{amihud}_{s,t}$: Amihud (2002) illiquidity measure for stock s in month t divided by its standard deviation in the previous 12 months; coefficient of −1 represents a 1sd reduction in price impact. $\texttt{spread}_{s,t}$: average bid-ask spread for stock s in month t divided by its standard deviation in the previous 12 months; coefficient of −1 represents a 1sd reduction in a stock's bid-ask spread. $\texttt{\#nhbr}_{s,t}$: number of neighbors to stock s in month t in thousands. Numbers in parentheses are standard errors clustered by stock. Statistical significance: $\star$ = 10%, $\star\star$ = 5%, and $\star\star\star$ = 1%. *Reads: "Stocks that have more neighbors in the ETF rebalancing network tend to realize more unpredictable ETF rebalancing demand and be more liquid as a result."*

References

Aaronson, S. (2013). *Quantum computing since Democritus*. Cambridge University Press.

Admati, A. (1985). A noisy rational-expectations equilibrium for multi-asset securities markets. *Econometrica*.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*.

Andrade, G., M. Mitchell, and E. Stafford (2001). New evidence and perspectives on mergers. *Journal of Economic Perspectives*.

Anton, M. and C. Polk (2014). Connected stocks. *Journal of Finance*.

Aumann, R. (1976). Agreeing to disagree. *Annals of Statistics*.

Azevedo, E. and E. Budish (2018). Strategy-proofness in the large. *Review of Economic Studies*.

Bai, Q., S. Bond, and B. Hatch (2015). The impact of leveraged and inverse ETFs on underlying real-estate returns. *Real Estate Economics*.

Barber, B. and T. Odean (2000). Trading is hazardous to your wealth: The common-stock investment performance of individual investors. *Journal of Finance*.

Barberis, N., A. Shleifer, and J. Wurgler (2005). Comovement. *Journal of Financial Economics*.

Baumol, W. and J. Benhabib (1989). Chaos: significance, mechanism, and economic applications. *Journal of Economic Perspectives*.

Ben-David, I., F. Franzoni, and R. Moussawi (2017). Do ETFs increase volatility? *Journal of Finance*.

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

Bernheim, D. and A. Rangel (2009). Beyond revealed preference: Choice-theoretic foundations for behavioral welfare economics. *Quarterly Journal of Economics*.

Bessembinder, H. (2015). Predictable ETF order flows and market quality. *Journal of Trading*.

Bessembinder, H., A. Carrion, L. Tuttle, and K. Venkataraman (2016). Liquidity, resiliency and market quality around predictable trades: Theory and evidence. *Journal of Financial Economics*.

Black, F. (1985). Noise. *Journal of Finance*.

Bollobás, B. (2001). *Random Graphs*. Cambridge University Press.

Brock, W. and C. Hommes (1997). A rational route to randomness. *Econometrica*.

Brown, D., S. Davies, and M. Ringgenberg (2016). ETF arbitrage and return predictability. Working Paper.

Budish, E. (2011). The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes. *Journal of Political Economy*.

Chang, Y., H. Hong, and I. Liskovich (2014). Regression discontinuity and the price effects of stock-market indexing. *Review of Financial Studies*.

Chinco, A. (2018). The madness of crowds and the likelihood of bubbles. Working Paper.

Cook, S. (1971). The complexity of theorem-proving procedures. In *Proceedings of the third annual ACM symposium on Theory of computing*.

De Long, B., A. Shleifer, L. Summers, and R. Waldmann (1990). Noise-trader risk in financial markets. *Journal of Political Economy*.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*.

Diamond, D. and P. Dybvig (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*.

Dow, J. and G. Gorton (1997). Noise trading, delegated portfolio management, and economic welfare. *Journal of Political Economy*.

Duffie, D., N. Gârleanu, and L. Pedersen (2005). Over-the-counter markets. *Econometrica*.

Erdos, P. and A. Rényi (1960). On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*.

Fortnow, L. (2017). *Golden Ticket: P, NP, and the Search for the Impossible*. Princeton University Press.

Frank, M. and T. Stengos (1989). Measuring the strangeness of gold and silver rates of return. *Review of Economic Studies*.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *Quarterly Journal of Economics*.

Garey, M. and D. Johnson (2002). *Computers and intractability*. WH Freeman New York.

Gennaioli, N. and A. Shleifer (2010). What comes to mind. *Quarterly Journal of Economics*.

Gill, J. (1977). Computational complexity of probabilistic turing machines. *SIAM Journal on Computing*.

Glosten, L. and P. Milgrom (1985). Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*.

Greenwood, R. and D. Thesmar (2011). Stock-price fragility. *Journal of Financial Economics*.

Gromb, D. and D. Vayanos (2010). Limits of arbitrage. *Annual Review of Financial Economics*.

Grossman, S. (1976). On the efficiency of competitive stock markets where trades have diverse information. *Journal of Finance*.

Grossman, S. and J. Stiglitz (1980). On the impossibility of informationally efficient markets. *American Economics Review*.

Hellwig, C. and L. Veldkamp (2009). Knowing what others know: Coordination motives in information acquisition. *Review of Economic Studies*.

Hellwig, M. (1980). On the aggregation of information in competitive markets. *Journal of Economic Theory*.

Hsieh, D. (1991). Chaos and nonlinear dynamics: application to financial markets. *Journal of Finance*.

Israeli, D., C. Lee, and S. Sridharan (2017). Is there a dark side to exchange-traded funds? an information perspective. *Review of Accounting Studies*.

Ivanov, I. and S. Lenkey (2018). Do leveraged ETFs really amplify late-day returns and volatility? *Journal of Financial Markets*.

Jackson, M. (2010). *Social and economic networks*. Princeton university press.

Karp, R. (1972). Reducibility among combinatorial problems. In *Complexity of Computer Computations*. Springer US.

Kyle, A. (1985). Continuous auctions and insider trading. *Econometrica*.

Madhavan, A. (2016). *Exchange-Traded Funds and the New Dynamics of Investing*. Oxford University Press.

Milgrom, P. and N. Stokey (1982). Information, trade, and common knowledge. *Journal of Economic Theory*.

Moore, C. and S. Mertens (2011). *The nature of computation*. OUP Oxford.

Mullainathan, S. (2002). A memory-based model of bounded rationality. *Quarterly Journal of Economics*.

Natarajan, B. (1995). Sparse approximate solutions to linear systems. *SIAM Journal on Computing*.

Newman, M., S. Strogatz, and D. Watts (2001). Random graphs with arbitrary degree distributions and their applications. *Physical Review E*.

Pagnotta, E. and T. Philippon (2018). Competing on speed. *Econometrica*.

Rosser, B. (1999). On the complexities of complex economic dynamics. *Journal of Economic Perspectives*.

Scheinkman, J. and B. LeBaron (1989). Nonlinear dynamics and stock returns. *Journal of Business*.

Shiller, R. (1984). Stock prices and social dynamics. *Brookings Papers on Economic Activity*.

Shleifer, A. (1986). Do demand curves for stocks slope down? *Journal of Finance*.

Shleifer, A. and L. Summers (1990). The noise-trader approach to finance. *Journal of Economic Perspectives*.

Shleifer, A. and R. Vishny (1997). The limits of arbitrage. *Journal of Finance*.

Shum, P., W. Hejazi, E. Haryanto, and A. Rodier (2015). Intraday share-price volatility and leveraged-ETF rebalancing. *Review of Finance*.

Sims, C. (2003). Implications of rational inattention. *Journal of Monetary Economics*.

Stein, J. (2009). Presidential address: Sophisticated investors and market efficiency. *Journal of Finance*.

Vayanos, D. and P. Woolley (2013). An institutional theory of momentum and reversal. *Review of Financial Studies*.

Verrecchia, R. (1982). Information acquisition in a noisy rational-expectations economy. *Econometrica*.

Wigderson, A. (1992). The complexity of graph connectivity. In *International Symposium on Mathematical Foundations of Computer Science*, pp. 112–132.

Wurgler, J. (2010). On the economic consequences of index-linked investing. In *Challenges to Business in the Twenty-First Century: The Way Forward*.

A Internet Appendix

Remark (Time Complexity)**.** Let Prob_1 and Prob_2 denote decision problems with instances of size S. Prob_1 is solvable in polynomial time if there is a solution algorithm that runs in $\mathrm{O}[S^k]$ steps for some $k > 0$. Whereas, Prob_2 requires exponential time if every solution algorithm requires $2^{\ell \cdot S}$ steps on at least one instance for some $\ell > 0$.

Decision problems with polynomial-time solutions are considered tractable while those that require exponential time are not. However, a polynomial-time solution for Prob_1 could require a $k = 10000$, and an exponential-time solution for Prob_2 could use an $\ell = 0.00001$. For these values of k and ℓ, Prob_2 would be easier to solve than Prob_1 on more reasonable instance sizes.

"If cases like this regularly arose in practice, then it would have turned out that we were using the wrong abstraction. But so far, it seems like we are using the right abstraction. Of the big problems solvable in polynomial time—matching, linear programming, primality testing, etc. . . —most of them really do have practical algorithms. And of the big problems that we think take exponential time—theorem-proving, circuit minimization, etc. . . —most of them really do not have practical algorithms. (Aaronson, 2013)" In short, when seen in this context, your first guess for both k and ℓ should be something like 1, 2, or 3.

Remark (Random Networks)**.** To make predictions about the likelihood of being affected by a rebalancing cascade, we assume a data-generating process for the market structure. A standard way to do this is to use a random-networks model (Jackson, 2010). The particular random-networks model we use dates back to Erdos and Rényi (1960). We chose this model because it is the simplest. Our main economic insight is about complexity not about random networks. Proposition 2.2b can be extended to other models with power-law and exponential edge distributions. See Newman et al. (2001) for more details.

Remark (Percolation Threshold)**.** The largest connected component of a directed graph is the largest set of nodes that are each connected to one another by a path. There is a sharp phase transition in the size of the largest connected component in an Erdös-Rényi model (Bollobás, 2001). When $\kappa < 1$, the size of the largest connected component remains finite as $S \to \infty$; whereas, when $\kappa > 1$, the largest connected component is infinitely large as $S \to \infty$. i.e., the largest connected component includes a finite fraction of infinitely many nodes. When $\kappa > 1$, the largest connected component is called the 'Giant Component'. For our purposes, this threshold implies the probability of asset Z being affected by a rebalancing cascade is vanishingly small when $\kappa < 1$.

Remark (Connectivity Threshold)**.** There is a similar phase transition in the existence of small connected components for the Erdös-Rényi random-networks model (Bollobás, 2001). When $\kappa < \log S$, the typical random network will contain many small connected components; whereas, when $\kappa > \log S$, the typical random network will contain only a giant component and nodes without any edges whatsoever. For our purposes, this threshold implies the probability stock Z is not affected by an index-fund rebalancing cascade starting somewhere else in the market is vanishingly small when $\kappa > \log S$.

Derivation (Equation 14)**.** Let $R_{s,t}$ denote the day-t return on stock s, and let $R_{b,t}$ denote the day-t return on the fth ETF's value-weighted benchmark index. Similarly, let $size_{s,t}$ denote the market capitalization of stock s on day t, and let $size_{b,t}$ denote the market capitalization of the fth ETF's benchmark index on day t. Then, the change in the fth ETF's value-weighted portfolio weight for stock s, $\Omega^{vw}_{f,s,t} = size_{s,t}/size_{b,t}$, is:

$$\begin{aligned}
\Delta\Omega^{vw}_{f,s,t} &= \frac{size_{s,t}}{size_{b,t}} - \frac{size_{s,t-1}}{size_{b,t-1}} \\
&= \frac{R_{s,t} \cdot size_{s,t-1}}{R_{b,t} \cdot size_{b,t-1}} - \frac{size_{s,t-1}}{size_{b,t-1}} \\
&= \frac{R_{s,t}}{R_{b,t}} \cdot \frac{size_{s,t-1}}{size_{b,t-1}} - \frac{size_{s,t-1}}{size_{b,t-1}} \\
&= \left(\frac{R_{s,t}}{R_{b,t}} - 1\right) \times \frac{size_{s,t-1}}{size_{b,t-1}} \\
&= \left(\frac{R_{s,t}}{R_{b,t}} - 1\right) \times \Omega_{f,s,t-1}
\end{aligned}$$

□

Discussion (Table A1)**.** Table A1 reports summary statistics for the number of M&A announcements each month during our sample period. These announcements represent the initial publicly observed shocks that trigger ETF rebalancing cascades in our empirical analysis. However, we emphasize that M&A announcements are only one kind of initial shock. Thus, the ETF rebalancing activity that we use analyze is only a small portion of the total noise generated by ETF rebalancing cascades.

Discussion (Table A2 and Figure A2)**.** Table A2 contains summary statistics describing above- and below-median stock Zs. This table shows that above-median stock Zs tend to be slightly different than below-median stock Zs on average. However, there is no reason to expect the size of this difference to increase immediately after an M&A announcement

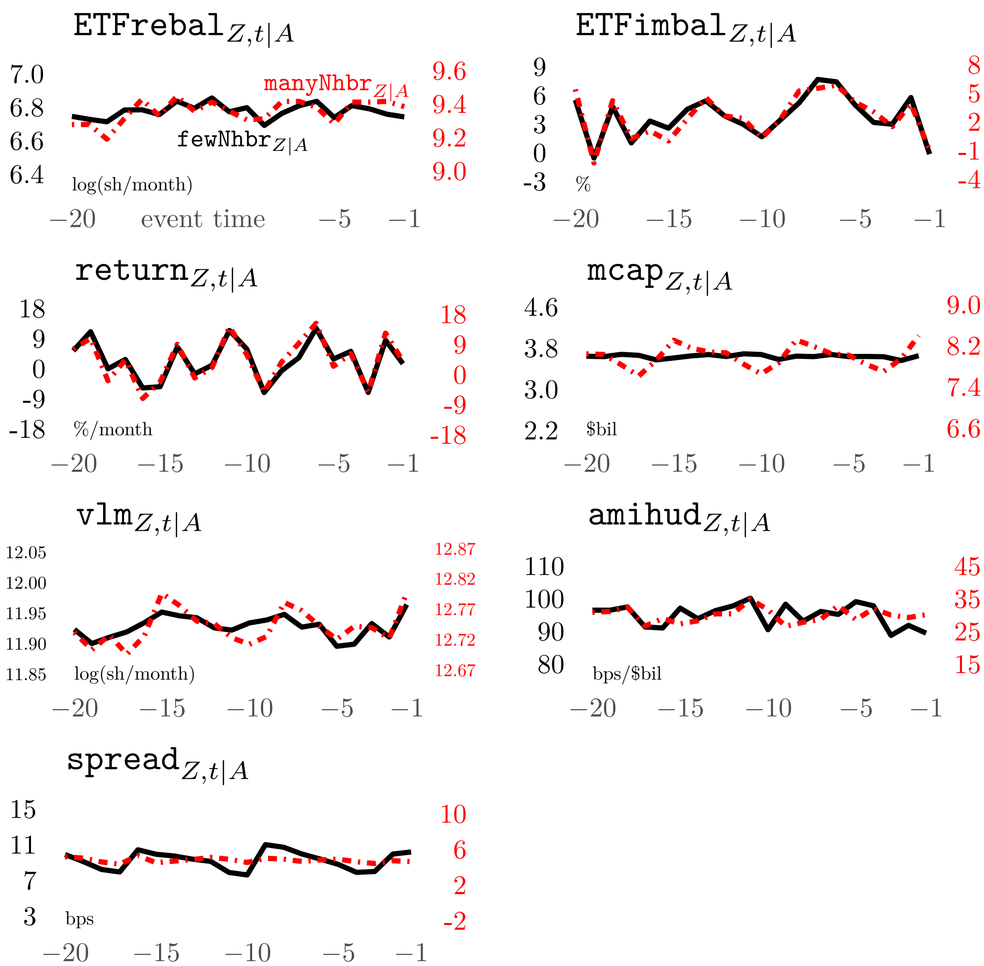


Figure A2. No Pre-Trends. Average ETF activity and characteristics of stock Zs during the 20 days prior to an M&A announcement about stock A. x-axis: event time with M&A announcement occurring on day 0. Red, Dashed: average value for stock Zs with an above-median number of neighbors; right y-axis. Black, Solid: average for stock Zs with a below-median number of neighbors; left y-axis. ETFrebal$_{Z,t|A}$: ETF rebalancing volume reported on a base-e logarithmic scale. ETFimbal$_{Z,t|A}$: ratio of signed ETF rebalancing volume to total ETF rebalancing volume in percent. return$_{Z,t|A}$: return in percent per month. mcap$_{Z,t|A}$: market capitalization in billions of dollars. vlm$_{Z,t|A}$: number of shares traded per month reported on a base-e logarithmic scale. amihud$_{Z,t|A}$: Amihud (2002) illiquidity measure over previous 20 days in basis points per \$1 million order. spread$_{Z,t|A}$: bid-ask spread as a fraction of the daily midpoint in basis points.

M&A Announcements

	Avg	Sd	25%	50%	75%
$\texttt{\#ancmt}_t$	14.3	6.2	10	14	18
$\texttt{\#stockZ}_t$	1.2	1.3	0.5	0.6	1.2

Table A1. Number of M&A announcements about publicly traded target firms each month. Data come from Thompson Financial. Sample: January 2011 to December 2017. There are 1199 total announcements in our sample. $\texttt{\#ancmt}_t$: number of announcements per month. $\texttt{\#stockZ}_t$: number of unrelated stock Zs for a given announcement in thousands.

about an unrelated stock A in the absence of ETF rebalancing cascades.

This parallel-trends assumption is something we can test in the data. For example, the upper-left panel of Figure A2 shows that, although above- and below-median stock Zs tend to have different amounts of ETF trading activity, this difference is constant in the run up to each M&A announcement. The remaining panels in this figure confirm that the same statement holds true for other stock Z characteristics, such as ETF order imbalance, returns, size, trading volume, and liquidity.

To re-iterate, we should expect to observe differences between above- and below-median stock Zs on average. After all, above-median stock Zs are always more likely to be hit by ETF rebalancing cascades. Our identification comes from the fact that these differences suddenly change in the wake of an unrelated M&A announcement.

Discussion (Table A3)**.** Table A3 reports the results of the second placebo test referenced on p. 32 of the main text. The first placebo test, whose results we report in column (5) of Table 3, checked to make sure that there was no difference between the ETF rebalancing volume of above- and below-median stock Zs following alternative announcement dates when no initial shock to stock A took place.

Now, we are going to use the right announcement date and instead restrict our attention to the rebalancing volume coming from the 30% of ETFs in our sample that are the least likely to rebalance each day. All of the ETFs in our sample rebalance more than once a quarter. But, some of the ETFs rebalance much more than others—once a day vs. twice a quarter. The ETFs that rebalance the least should also be the least likely to transmit the effects of an ETF rebalancing cascade.

Consistent with this prediction, Table A3 shows that, when we restrict our attention to these infrequent rebalancers, above-median stock Zs have ETF rebalancing volume that is

Summary Statistics, Stock Z

Panel A. ETF Activity

	All		Number of Neighbors Many		Few	
	Avg	Sd	Avg	Sd	Avg	Sd
$\texttt{\#ETFholding}_{Z\|A}$	16.2	18.9	24.4	21.3	8.7	12.4
$\texttt{\#nhbr}_{Z\|A}$	2.5	1.5	3.1	1.2	1.9	1.5
$\texttt{ETFvlm}_{Z\|A}$	7.8	4.0	9.3	3.6	6.5	3.9
$\texttt{ETFrebal}_{Z\|A}$	7.9	3.7	9.2	3.4	6.7	3.6
$\texttt{ETFimbal}_{Z\|A}$	3.3	26.4	2.5	24.4	4.0	28.2

Panel B. Characteristics

	All		Number of Neighbors Many		Few	
	Avg	Sd	Avg	Sd	Avg	Sd
$\texttt{return}_{Z\|A}$	4.4	89.1	5.4	87.1	3.4	90.8
$\texttt{mcap}_{Z\|A}$	5.7	19.9	7.9	24.9	3.6	13.1
$\texttt{vlm}_{Z\|A}$	12.2	2.3	12.7	2.1	11.8	2.5
$\texttt{amihud}_{Z\|A}$	73.9	636.6	41.3	599.7	103.9	667.4
$\texttt{spread}_{Z\|A}$	9.6	54.3	7.7	52.1	11.3	56.1

Table A2. Summary statistics for the set of stock Zs unrelated to each M&A announcement. An observation is the average value for a stock Z during the 20 days leading up at an M&A announcement. Data on M&A announcements with publicly traded target firms comes from Thompson Financial. Stock-market data comes from CRSP. Sample: January 1st, 2011 to December 31st, 2017. All: all stock Zs that are unrelated to each M&A announcement. Many Neighbors: stock Zs with an above-median number of neighbors for a particular M&A announcement. Few Neighbors: stock Zs with a below-median number of neighbors for a particular M&A announcement. **Panel A.** ETF activity for each stock. $\texttt{\#ETFholding}_{Z|A}$: number of ETFs that hold a stock. $\texttt{\#nhbr}_{Z|A}$: number of neighboring stocks in the ETF rebalancing network in thousands. $\texttt{ETFvlm}_{Z|A}$: ETF trading volume each month on a base-e logarithmic scale. $\texttt{ETFrebal}_{Z|A}$: ETF rebalancing volume each month on a base-e logarithmic scale. $\texttt{ETFimbal}_{Z|A}$: signed ETF rebalancing volume divided by total ETF rebalancing volume reported in percent. **Panel B.** Characteristics of each stock. $\texttt{return}_{Z|A}$: return in percent per month. $\texttt{mcap}_{Z|A}$: market capitalization in billions of dollars. $\texttt{vlm}_{Z|A}$: total trading volume each month on a base-e logarithmic scale. $\texttt{amihud}_{Z|A}$: Amihud (2002) illiquidity measure computed daily within a given month in basis points per \$1 million order. $\texttt{spread}_{Z|A}$: average bid-ask spread as a fraction of midpoint during month in basis points.

statistically indistinguishable from below-median stock Zs in the five days after an M&A announcement. These findings further support the claim that our results are due to the interaction of ETF rebalancing rules and not some as-yet unknown omitted variable.

Discussion (Table A4)**.** Table A4 reports the change in ETF rebalancing volume for stock A (distance = 0), stock A's sister stocks (distance = 1), stock Zs that are at least twice

ETF Rebalancing Volume, Stock Z Least-Active ETFs

Dependent Variable: $\texttt{ETFrebal}_{Z,t|A}$

	(1)	(2)
$\texttt{afterAncmt}_{t\vert A} \times \texttt{manyNhbr}_{Z\vert A}$	0.24 (0.37)	0.11 (0.37)
$\texttt{afterAncmt}_{t\vert A}$	1.92*** (0.27)	1.51*** (0.26)
$\texttt{vlm}_{Z,t\vert A}$		26.60*** (0.55)
Announcement×#nhbr FE	Y	Y
Stock-Z FE	Y	Y
R^2	80.0%	80.2%
Observations	20,312,414	

Table A3. Effect of initial M&A announcement about stock A on the ETF rebalancing volume of an unrelated stock Z when looking only at ETF rebalancing volume for the 30% of ETFs that rebalance their positions the least frequently. Sample: January 1st, 2011 to December 31st, 2017. Each column presents the results of a separate regression using daily data on the 26-day window surrounding each M&A announcement, $\{t_A - 20, \ldots, t_A - 1, t_A, t_A + 1, \ldots, t_A + 5\}$. $\texttt{ETFrebal}_{Z,t|A}$: dependent variable is ETF rebalancing volume for stock Z on date t following M&A announcement about stock A on a base-e logarithmic scale; coefficient of +1 indicates a 1% per day increase in a stock's ETF rebalancing volume. $\texttt{afterAncmt}_{t|A}$: indicator variable that is one during the five days following an M&A announcement about stock A. $\texttt{manyNhbr}_{Z|A}$: indicator variable that is one if stock Z has an above-median number of neighbors in the ETF rebalancing network relative to the M&A announcement about stock A. $\texttt{vlm}_{Z,t|A}$: total trading volume for stock Z on a given day reported on a base-e logarithmic scale. Numbers in parentheses are standard errors clustered by stock Z. Statistical significance: $\star$ = 10%, $\star\star$ = 5%, and $\star\star\star$ = 1%. *Reads: "When we restrict our attention to the 30% of ETFs that are the least likely to rebalance each day, stock Zs with many neighbors have ETF rebalancing volume that is statistically indistinguishable from stock Zs with few neighbors in the five days after an M&A announcement."*

removed (distance ≥ 2), and stock Zs that are more than twice removed (distance ≥ 3). As already evident from Tables 2 and 3, comparing column (1) to columns (2)-(4) confirms that ETF rebalancing volume is much higher for stock A than for any other stock in the network. This result is consistent with the observation that not all ETF rebalancing rules look like the ones we model in our theoretical analysis.

However, columns (2)-(4) in Table A4 show that, after the immediate drop from stock A to any other stock, the increase in ETF rebalancing volume does not decay as we move from 1 to ≥ 3 links away in the ETF rebalancing network on the day of the announcement. This evidence is consistent with the importance of end-of-day ETF rebalancing and of the one-for-one rebalancing decisions we study in our theoretical framework.

ETF Rebalancing Volume, Stock Z
Effect of Distance

Dependent Variable: $\texttt{ETFrebal}_{s,t}$				
Distance to stock A:	0	1	≥ 2	≥ 3
	(1)	(2)	(3)	(4)
$1_{\{t=t_A\}}$	154.29⋆⋆⋆ (6.73)	2.89⋆⋆⋆ (0.07)	2.14⋆⋆⋆ (0.19)	2.70⋆⋆⋆ (0.25)
Month-Year FE	Y	Y	Y	Y
Stock FE	Y	Y	Y	Y
R^2	78.5%	66.5%	59.4%	75.0%
Observations	15,459	44,853,789	23,264,687	16,021,280

Table A4. Effect of initial M&A announcement on the ETF rebalancing volume of stocks at various distances from stock A. Sample: January 1st, 2011 to December 31st, 2017. $\texttt{ETFrebal}_{Z,t|A}$: ETF rebalancing volume on date t following M&A announcement about stock A on a base-e logarithmic scale; coefficient of +1 indicates a 1% per day increase in a stock's ETF rebalancing volume. Each column presents the results of a separate regression using daily data on the 26-day window surrounding each M&A announcement, $\{t_A - 20, \ldots, t_A - 1, t_A, t_A + 1, \ldots, t_A + 5\}$. Column (1) reports the change in ETF rebalancing volume for stock A (distance = 0); column (2) reports the change for stock A's sister stocks (distance = 1); column (3) reports the change for stock Zs that are at least twice removed (distance ≥ 2); and, column (4) reports the change for stock Zs that are at least thrice removed (distance ≥ 3). $1_{\{t=t_A\}}$: indicator variable that is one if an observation was on stock A's M&A announcement date. Numbers in parentheses are standard errors clustered by stock Z. Statistical significance: ⋆ = 10%, ⋆⋆ = 5%, and ⋆⋆⋆ = 1%. *Reads: "Alternative rebalancing rules mute ETF rebalancing in stock A's sister stocks. But, the one-for-one rebalancing rules we model in our theoretical framework are prevalent enough that ETF rebalancing volume does not decay much with distance once a cascade has begun."*