The Sound of Many Funds Rebalancing*

Alex Chinco† and Vyacheslav Fos‡

February 8, 2018

Abstract

This paper proposes that complexity generates noise in financial markets. A stock's demand might appear random, not because individual investors are behaving randomly, but because it's too computationally complex to predict how a wide variety of simple, deterministic, trading rules will interact with one another—even if you yourself are fully rational. There are two parts to our analysis. First, we illustrate how complexity can generate noise by modeling a particular kind of trading-rule interaction: index-fund rebalancing cascades. An initial shock to stock A causes an index fund to buy stock A and sell stock B, which then causes a second fund following a different benchmark to sell stock B and buy stock C, which then causes a third fund following yet another benchmark to... Although it's easy to predict *if* this index-fund rebalancing cascade will eventually affect the demand for an unrelated stock Z, predicting *how* stock Z will be affected (buy? or sell?) is computationally intractable. Second, we give empirical evidence that complexity actually does generate noise in real-world financial markets by analyzing the end-of-day holdings of exchange-traded funds (ETFs). We show that ETF rebalancing cascades transmit economically large demand shocks, which are also statistically unpredictable. And, we document market participants behaving as if these demand shocks are noise.

JEL Classification. G00, G02, G14

Keywords. Noise, Indexing, Thresholds

---

*We thank Kerry Back, Nick Barberis, Zahi Ben-David, James Choi, Adam Clark-Joseph, Tony Cookson, Xavier Gabaix, Itay Goldstein, Sam Hartzmark, Ralph Koijen, Pete Kyle, Chris Parsons, Jeff Pontiff, and Brian Weller, as well as seminar participants at CalTech, Colorado, Illinois, Maryland, Yale, 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, and the FRA Conference for extremely helpful comments and suggestions.

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

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

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

1 Introduction

Imagine you're a trader who's just discovered that stock Z is under-priced. In a market without noise, there's no way for you to take advantage of this discovery. The moment you try to buy a share, other traders will immediately realize that you must have uncovered some good news. And, you won't find anyone willing to sell you a share at the old price (Aumann, 1976; Milgrom and Stokey, 1982).

Noise pulls the rug out from under this no-trade theorem. In a market with noise, someone may always be trying to buy or sell stock Z for erratic non-fundamental reasons. So, when you try to buy a share, other traders won't immediately realize that you've uncovered good news since your buy order could just be some more random noise. It's this plausible cover story that allows you to both trade on and profit from your discovery. It's this plausible cover story that Fischer Black was referring to when he wrote that "noise makes financial markets possible".

But, where exactly does this all-important noise come from? Who generates it? And, what are their erratic non-fundamental reasons for trading?

The standard answer to these questions is that i) noise comes from individual investors, and that ii) their demand looks erratic and unrelated to fundamentals because individual investors are just plain bad traders. These are the standard answers for a reason. Not only do individual investors suffer from all sorts of behavioral biases when they trade (Barberis and Thaler, 2003) but they trade far too often (Barber and Odean, 2000). So, it's clear that individual investors can generate noise.

But, are they the only source? It seems unlikely. The importance of individual investors has steadily declined over the past few decades. While individual investors held 47.9% of all U.S. equity in 1980, this percentage was down to only 21.5% by 2007 (French, 2008). And, in June 2017 JP Morgan strategists reported that only around "10% of trading is done by traditional, 'discretionary' traders, as opposed to systematic rules-based ones."[1] However, this steady downward trend in the importance of individual investors has not been accompanied by a drop in trading volume.

Motivated by this explanatory gap, we propose an alternative noise-generating mechanism based on computational complexity. A stock's demand might appear random, not because individual investors are behaving randomly, but because it's too computationally complex to predict how a wide variety of simple, deterministic, trading rules will interact with one another—even if you yourself are fully rational. There

[1]Financial Times. 6/14/2017. *Not Your Father's Market: Tech Tantrum Shows How U.S. Equities Trading Has Changed.*

are two parts to our analysis. First, we show theoretically how computational complexity can generate noise by modeling a particular kind of trading-rule interaction: index-fund rebalancing cascades. Then, we give empirical evidence that index-fund rebalancing cascades actually generate noise in real-world financial markets using data on the end-of-day holdings of exchange-traded funds (ETFs).

Theoretical Model. As individual investors have shrunk in importance, "passive investing—indexing—has become popular as an alternative to active investment management" and "active managers... have become more index-like in their investing (Stambaugh, 2014)." However, these new 'index-like' funds have not been created in Jack Bogle's image. Many choose their holdings "based on custom criteria" that involve threshold-based rules.[2] For instance, the PowerShares S&P 500 Low-Volatility ETF [`SPLV`] tracks the lowest-volatility quintile of S&P 500 stocks. This fund uses a threshold-based 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` has to exit its position in one stock and build a new position in another, affecting each stock in equal-but-opposite ways. The price of the stock being added will rise while the price of the 'stock formerly known as 100th' will fall.

We begin our analysis by presenting a model where, because there are so many index funds tracking so many different threshold-based benchmarks, a small change in stock A's price can cause one index fund to buy stock A and sell stock B, which can then cause a second index fund following a different threshold-based benchmark to sell stock B and buy stock C, which can then cause... Our main theoretical result is that, although it's possible to determine *if* a stock will be affected by one of these index-fund rebalancing cascades, the problem of determining *how* the stock will be affected (buy? or sell?) is computationally intractable. In fact, it's NP hard. Thus, index-fund rebalancing cascades can generate seemingly random demand shocks even though each index fund involved in the cascade is following a completely deterministic trading rule. In other words, index-fund rebalancing cascades generate noise in a way that does not require traders to suffer from behavioral biases or make cognitive errors.

Rebalancing Cascades. Our theoretical model shows how index-fund rebalancing cascades are able to generate seemingly random demand shocks. But, is there any evidence that they are actually doing this in real-world financial markets? To answer this question, we study end-of-day ETF holdings using data from ETF Global, which covers every trading day from January 2010 to December 2015. We focus our attention

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

on ETFs that rebalance daily. So, when you look at our results, you should be thinking about the PowerShares S&P 500 Low-Volatility ETF [SPLV] rather than the SPDR S&P 500 ETF [SPY]. ETFs that rebalance daily are smaller than funds that track broad value-weighted market indexes, such as SPY. But, their rebalancing activity matters because these funds tend to do the bulk of their trading during the final 20-to-30 minutes of the trading day.[3] We also net-out changes in ETF holdings due to creations and redemptions. These trades are executed as in-kind transfers for tax reasons (Madhavan, 2016) and so can't contribute to index-fund rebalancing cascades.

Here's how we structure our tests. Our theoretical model studies index-fund rebalancing cascades that stem from an initial shock. So, we study ETF rebalancing cascades that stem from an initial M&A announcement, referring to the target of an M&A announcement as stock A. Our data on M&A announcements comes from Thomson Financial. M&A deals are a natural choice for the initial shocks 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)." While M&A targets are not randomly chosen, the exact date of the announcement (Tuesday? Wednesday? or Thursday?) may as well be.

Following each stock A's announcement as an M&A target, we examine the set of unrelated stock Zs. For stock A and stock Z to be unrelated, they have to be twice removed in the network of ETF holdings at the time of the M&A announcement. Stock Z can't have been recently held by any ETF that also held stock A. And, if stock A and stock B both belong to the same benchmark, then stock Z can't have been recently held by any ETF that also held stock B. In other words, the chain has to be $A \to B \to C \to Z$ or longer. Because there are smart-beta ETFs focusing on things like large-cap, value, and industry-specific benchmarks, this unrelatedness criteria also implies that stock Z doesn't share any well-known characteristics such as size, book-to-market, or industry with stock A.

Our theoretical model predicts that, all else equal, a stock Z that's on the cusp of more ETF rebalancing thresholds is more likely to be hit by an ETF rebalancing cascade. So, we split the set of stock Zs for each M&A announcement into two subsets: those that are on the cusp of rebalancing for an above-median number of ETFs, and those that aren't. Consistent with our theoretical prediction, we find that ETF rebalancing volume grows by 169% more for the above-median group of stock Zs than for the below-median group in the 5 days immediately following an M&A

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

announcement. And, we show that this increase is no more likely to be made up of buy orders than of sell orders. Taken together, these results suggest that it's possible to predict *if* stock Z will be affected by an ETF rebalancing cascade but not *how* stock Z will be affected. What's more, because the same stock Z can be above-median relative to one M&A target but below-median relative to another, our results can't be attributed to unobserved characteristics of stock Z.

Market Reaction. To show that ETF rebalancing cascades are actually generating noise, however, we need to do more than just give statistical evidence that they are unpredictable. To be convincing, we need to show that market participants treat the resulting demand shocks as noise, too. The natural way to investigate whether market participants are treating the demand coming from ETF rebalancing cascades as noise is to look for differences in liquidity. As highlighted in the opening paragraphs, if there is more demand noise, then a randomly selected trade is less likely to be coming from an informed trader. So, we look for otherwise identical stock Zs and ask: 'Is stock Z_1 more liquid than stock Z_2 at times when stock Z_1 happens to be more susceptible to ETF rebalancing cascades than stock Z_2?' The data reveals that the answer is 'Yes'. Above-median stock Zs are much more liquid than the below-median stock Zs suggesting that market participants treat the demand shocks coming from ETF rebalancing cascades as noise.

Broader Implications. John Maynard Keynes (1921) pointed out that, because a daily national census is logistically impossible, the answer to the question 'Is the population of France an even or an odd number?' is effectively a coin toss. So, economists had intuited a connection between apparent randomness and computational complexity long before ETFs arrived on the scene. But, in the past, this intuitive connection his always been just that—an intuition. The goal of our theoretical model is to make the connection concrete.

By showing precisely why it's computationally intractable to predict ETF rebalancing cascades even if you yourself are fully rational, we make it possible for researchers to identify other situations where the same logic holds. For example, our theoretical model also applies to any other group of funds following a wide variety of threshold-based rebalancing rules. Think about quantitative hedge funds following strategies of the form 'Buy the top 30% and sell the bottom 30% of stocks when sorting on X' (Khandani and Lo, 2007). Or, consider pension funds with strict portfolio mandates of the form '15% of our assets will be held in alternative investments' (Pennacchi and Rastad, 2011).

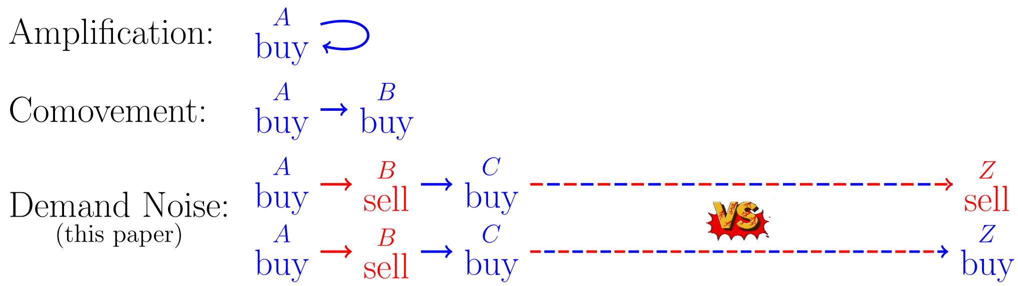


Figure 1: How This Paper Is Different. *Papers on index-linked investing fall into two groups. The first studies how trading due to index inclusion can amplify an initial shock to stock A (Row 1). The second studies how stock A's returns suddenly co-move with stock B's returns as soon as stock A gets added to stock B's index (Row 2). By contrast, this paper focuses on the unpredictable consequences of stock A's index inclusion, not for stock A or for stock B, but for seemingly unrelated stock Zs (Rows 3 and 4).*

1.1 Related Literature

This paper connects to three main strands of literature.

Noise. The problem we study is motivated by the central role that noise plays in information-based asset pricing models (Grossman and Stiglitz, 1980; Hellwig, 1980; Admati, 1985; Kyle, 1985) and limits-to-arbitrage models (Shleifer and Summers, 1990; Shleifer and Vishny, 1997; Gromb and Vayanos, 2010). We propose an explanation for noise that does not rely on individual investors behaving randomly.

Indexing. Our paper also relates to work on index-linked investing (Wurgler, 2010). Some of these papers study how index inclusion can amplify an initial shock to stock A. For instance, Chang et al. (2014) shows how getting added to the Russell 2000 results in further price increases. For examples involving ETFs, see Ben-David et al. (2017), Brown et al. (2016), and Israeli et al. (2017). Other papers study how stock A's returns suddenly co-move with stock B's returns as soon as stock A gets added to stock B's index. For instance, Barberis et al. (2005) shows that a stock's beta with the S&P 500 jumps sharply after index inclusion. For other examples, see Greenwood and Thesmar (2011), Vayanos and Woolley (2013), and Anton and Polk (2014). By contrast, we focus on the unpredictable consequences of stock A's index inclusion, not for stocks A or B, but for seemingly unrelated stock Zs.

Thresholds. Finally, people use threshold-based rules to make all sorts of decisions (Gabaix, 2014). The literature on heuristic decision making typically measures the cost of using a heuristic rule in *expected*-utility loss (Bernheim and Rangel, 2009). Whereas, we look at how simple decision rules can affect demand *volatility*.

2 Theoretical Model

Because there are so many index funds tracking so many different benchmarks, a small change in stock A's characteristics can cause one index fund to buy stock A and sell stock B, which can then cause a second index fund following a different benchmark to sell stock B and buy stock C, which can then cause... This section presents a model showing that, while it's possible to determine *if* a stock will be affected by one of these index-fund rebalancing cascades, predicting *how* the stock will be affected (buy? or sell?) is an NP-hard problem. As a result, demand shocks coming from index-fund rebalancing cascades are effectively random even though each index fund involved in a cascade is following a simple, completely deterministic, rebalancing rule.

2.1 Market Structure

Here's how we model index funds transmitting an initial shock from stock A to stock B, and then from stock B to stock C, and then from stock C to stock D, and so on.

Network. Imagine a market with a set of stocks $\mathsf{S} = \{1, 2, \ldots, S\}$. Index-fund rebalancing rules define a network over these stocks. There is an edge from stock s to stock s', not if they both currently belong to the same benchmark, but rather if a shock to stock s would cause an index fund to swap its position in stock s for a new position in stock s'. If a positive shock to stock s would cause some fund to sell stock s' and buy stock s, then stock s' is a negative neighbor to stock s:

$$\mathsf{N}_s^- = \{s' \in \mathsf{S} \,|\, \text{positive shock to } s \Rightarrow \text{negative shock to } s'\} \tag{1}$$

Whereas, if a negative shock to stock s would cause some fund to buy stock s' and sell stock s, then stock s' is a positive neighbor of stock s:

$$\mathsf{N}_s^+ = \{s' \in \mathsf{S} \,|\, \text{negative shock to } s \Rightarrow \text{positive shock to } s'\} \tag{2}$$

The market structure is the set of positive and negative neighbors for each stock, $\mathsf{M} = \{(\mathsf{N}_1^+ \| \mathsf{N}_1^-), (\mathsf{N}_2^+ \| \mathsf{N}_2^-), \ldots, (\mathsf{N}_S^+ \| \mathsf{N}_S^-)\}$.

Distortion. This network of rebalancing rules propagates shocks through the market in discrete rounds, which are indexed by $t = 0, 1, 2, \ldots$ For this to happen, index-fund rebalancing decisions must have the potential to distort stock characteristics. If one fund decides to sell stock s, then this decision must have the potential to change stock s in a way that causes a second fund to rebalance, too. In other words, it's important that demand curves slope down (Shleifer, 1986).

This assumption is consistent both with trader descriptions of index-fund rebalancing and with current academic research (Ben-David et al., 2017). More and more people are talking about how ETF rebalancing "influences trading in individual stocks."[4] And, there's a lot of overlap between index-fund portfolios. The same stock is often involved in numerous ETF benchmarks, such as "active beta, momentum, dividend growth, deep value, quality, and total earnings."[5]

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

$$x_{s,t} \in \{-1,\, 0, 1\} \qquad\qquad \Delta x_{s,t} = x_{s,t} - x_{s,t-1} \tag{3}$$

If $(x_{s,t}, \Delta x_{s,t}) = (1,1)$, then stock s has realized a positive shock because some fund built a new position in stock s. If $(x_{s,t}, \Delta x_{s,t}) = (-1,\, -1)$, then the opposite outcome has taken place. Stock s has realized a negative shock because some fund exited an existing position in stock s. To emphasize that index-fund rebalancing decisions can affect more than just a stock's price, we refer to changes in stock 'characteristics'.

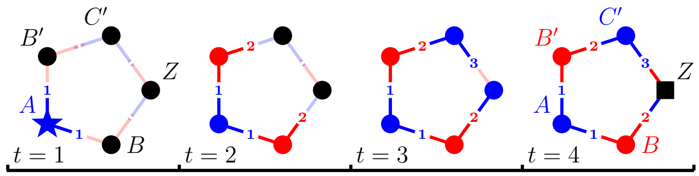
Propagation. Because we want to illustrate how computational complexity can generate seemingly random demand shocks even in the absence of any random behavior on the part of individual investors, we model how index-fund rebalancing decisions propagate shocks through the market as a mechanical 3-step process. STEP 1 involves identifying the set of stocks that will be affected at time $(t+1)$ by index funds' rebalancing decisions at time t:

$$\mathsf{Out}^{+}_{s,t} = \begin{cases} \big\{ s' \in \mathsf{N}^{+}_{s} \mid s \notin \mathsf{Out}^{-}_{s',t-1} \big\} & \text{if } (x_{s,t}, \Delta x_{s,t}) = (-1,\, -1) \\ \emptyset & \text{otherwise} \end{cases} \tag{4a}$$

$$\mathsf{Out}^{-}_{s,t} = \begin{cases} \big\{ s' \in \mathsf{N}^{-}_{s} \mid s \notin \mathsf{Out}^{+}_{s',t-1} \big\} & \text{if } (x_{s,t}, \Delta x_{s,t}) = (1, 1) \\ \emptyset & \text{otherwise} \end{cases} \tag{4b}$$

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

[4] Bloomberg. 4/10/2015. *Tail Can Wag Dog When ETFs Influence Single Stocks, Goldman Says.*
[5] Financial Times. 10/7/2017. *On The Perverse Economic Effects Created by ETFs.*



	$\mathsf{N}_s^+ \; \mathsf{N}_s^-$	$t=1$	$t=2$	$t=3$	$t=4$
A	$\| B, B'$	$+1$	$+1$	$+1$	$+1$
B	$A, Z \|$	0	-1	-1	-1
B'	$A, C' \|$	0	-1	-1	-1
C'	$\| B', Z$	0	0	$+1$	$+1$
Z	$C' \| B$	0	0	$+1$	$0 = \text{Effect}_{\mathsf{M},4}(\mathsf{A}, Z)$

$\text{Cascade}_{\mathsf{M},4}(\mathsf{A})$

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

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

$$\mathsf{In}^+_{s,t+1} = \{ s' \in \mathsf{S} \mid s \in \mathsf{Out}^+_{s',t} \} \tag{5a}$$

$$\mathsf{In}^-_{s,t+1} = \{ s' \in \mathsf{S} \mid s \in \mathsf{Out}^-_{s',t} \} \tag{5b}$$

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

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

$$u_{s,t+1} = 1_{\{|\mathsf{In}^+_{s,t+1}| > |\mathsf{In}^-_{s,t+1}|\}} - 1_{\{|\mathsf{In}^+_{s,t+1}| < |\mathsf{In}^-_{s,t+1}|\}} \tag{6a}$$

$$x_{s,t+1} = \text{Sign}[x_{s,t} + u_{s,t+1}] \tag{6b}$$

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

s, then it will realize a negative shock.

Cascades. An index-fund rebalancing cascade starts in round $t = 0$ with all stocks at their default levels:

$$(x_{s,0}, \Delta x_{s,0}) = (0, 0) \tag{7}$$

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

$$(x_{s,1}, \Delta x_{s,1}) = (1, 1) \qquad\qquad \text{for each } s \in \mathsf{A} \tag{8}$$

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

Following this initial shock, an index-fund rebalancing cascade is just the iteration of the 3-step updating procedure until a time limit $T \in \{1, 2, \ldots\}$ has been reached:

$$
\begin{array}{ll}
 & \textbf{function } \text{Cascade}_{\mathsf{M},T}(\mathsf{A})\textbf{:} \\
 & \quad t \leftarrow 0 \\
 & \quad \textbf{for all } (s \in \mathsf{A})\textbf{:} \\
 & \quad\quad (x_s, \Delta x_s) \leftarrow (1,1) \\
 & \quad \textbf{while } (t \leq T)\textbf{:} \\
 & \quad\quad \textbf{for all } (s \in \mathsf{S})\textbf{:} \\
\text{Step 1:} & \quad\quad\quad (\mathsf{Out}_s^+, \mathsf{Out}_s^-) \leftarrow \text{Update}[(\mathsf{Out}_s^+, \mathsf{Out}_s^-)|(x_s, \Delta x_s)] \\
 & \quad\quad \textbf{for all } (s \in \mathsf{S})\textbf{:} \\
\text{Step 2:} & \quad\quad\quad (\mathsf{In}_s^+, \mathsf{In}_s^-) \leftarrow \text{Update}[(\mathsf{In}_s^+, \mathsf{In}_s^-)] \\
\text{Step 3:} & \quad\quad\quad (x_s, \Delta x_s) \leftarrow \text{Update}[(x_s, \Delta x_s)] \\
 & \quad\quad t \leftarrow t + 1 \\
 & \quad \textbf{return } [\, x_1 \ x_2 \ \cdots \ x_S \,]
\end{array}
\tag{9}
$$

An index-fund rebalancing cascade's effect on stock Z, $\text{Effect}_{\mathsf{M},T}(\mathsf{A}, Z)$, is the Zth element of the output from $\text{Cascade}_{\mathsf{M},T}(\mathsf{A})$. Notice that how description of an index-fund rebalancing cascade suggests a second interpretation for the symbol M. M is not just a description of index-fund rebalancing rules. It's also a description of a machine that computes the effects of index-fund rebalancing cascades.

An Example. Figure 2 shows an example of an index-fund rebalancing cascade involving 5 stocks that starts with a positive shock to stock A. At time $t = 3$, the cascade delivers a positive shock to stock Z, $\text{Effect}_{\mathsf{M},3}(\{A\}, Z) = +1$. But then, at time $t = 4$, a second branch of the cascade hits stock 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 index-fund rebalancing cascades in the following two subsections. First, is there any way for an index-fund rebalancing cascade that starts at stock A to effect stock Z? Second, suppose there is. What will be the net effect of the rebalancing cascade on stock Z? In the next two subsections, we're going to investigate the computational complexity of answering each of these questions.

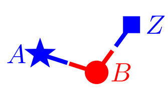
2.2 'If?' Problem

How hard is it to figure out whether an index-fund rebalancing cascade triggered by an initial shock to stock A might eventually affect the demand for stock Z?

Decision Problem. Solving this decision problem means finding at least one path connecting a particular stock A to stock Z. A K-path connecting stock A to stock Z is a sequence of K stocks $\{s_1, \ldots, s_K\}$ such that the first stock is stock A, the last stock is stock Z, and

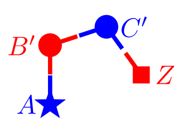
$$s_k \in \begin{cases} \mathsf{N}^+_{s_{k-1}} & k \text{ odd} \\ \mathsf{N}^-_{s_{k-1}} & k \text{ even} \end{cases} \qquad \text{for all } k \in \{2, \ldots, K\} \tag{10}$$

For example, in Figure 2, there are two different paths from stock A to stock Z. One travels from stock A to stock B to stock Z:



$$\{\underset{\text{Stock } A}{(\emptyset \| \{B\})}, \rightarrow \underset{\text{Stock } B}{(\{A,Z\} \| \emptyset)}, \rightarrow \underset{\text{Stock } Z}{(\emptyset \| \{B\})}\} \tag{11}$$

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



$$\{\underset{\text{Stock } A}{(\emptyset \| \{B'\})}, \rightarrow \underset{\text{Stock } B'}{(\{A,C'\} \| \emptyset)}, \rightarrow \underset{\text{Stock } C'}{(\emptyset \| \{B',Z\})}, \rightarrow \underset{\text{Stock } Z}{(\{C'\} \| \emptyset)}\} \tag{12}$$

If such a path exists, then it's possible that an index-fund rebalancing cascade triggered by an initial shock to stock A might affect the demand for stock Z.

Below we give a formal definition of the 'If?' problem.

Problem 2.2a (If).

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

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

If *Complexity.* Problems with polynomial-time solutions are considered "tractable problems" that "can be solved in a reasonable amount of time (Moore and Mertens, 2011)." And, the proposition below shows that If can be solved in polynomial time. So, it's easy to determine which stocks have the potential to trigger an index-fund rebalancing cascade that might affect stock 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 $y \geq y_0$. And, 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 stocks 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.

Predicting If. The computational tractability of If also means that you can make useful predictions about the size of $\hat{\mathsf{S}}$ for a given stock Z. To illustrate, suppose that for any pair of stocks $(s, s') \in \mathsf{S}^2$, stock s' is chosen as a positive neighbor to stock 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 stock, $N_s^+ = |\mathsf{N}_s^+|$, obeys a Poisson distribution as $S \to \infty$ (Erdos and Rényi, 1960)

$$N_s^+ \sim \mathrm{Poisson}(\kappa, S) \tag{13}$$

which implies that the typical stock has $\mathrm{E}[N_s^+] = \kappa$ positive neighbors. Thus, if $\kappa \approx 0$, then the market will be fragmented with most stocks having no neighbors; whereas, if $\kappa \approx \log(S)$, then the market will be densely connected with each stock on the cusp of rebalancing for many different funds.

The proposition below shows that it's easy to predict how many stocks are connected to stock Z just by counting the number of neighbors for stock Z.

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

$$\hat{\mathrm{S}}_{\max}(Z, \mathsf{M}, T) = \max_{\hat{\mathsf{S}} \in 2^{\mathsf{S}}} \left\{ \, |\hat{\mathsf{S}}| \ \ s.t. \ (Z, \mathsf{M}, T, \hat{\mathsf{S}}) \in \mathsf{If} \, \right\} \tag{14}$$

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

Put differently, stocks with more neighbors are more likely to be affected by index-fund rebalancing cascades. And, you can infer this property about stock Z without having to trace out each individual path that a rebalancing cascade might take. We will make use of this fact in our empirical analysis below.

2.3 'How?' Problem

Although it's easy to predict *if* a stock is likely to be affected by an index-fund rebalancing cascade, predicting *how* a stock will be affected is computationally intractable.

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 3. Each row depicts a single market with $S = 25$ stocks and is broken up into 3 panels. Here's the exercise we have in mind. First, examine the left panel in each row, which depicts the index-fund rebalancing rules that define each market. Then, ask yourself: i) 'Will stock Z, which is denoted by the large black square with a question mark in it, be affected by an index-fund rebalancing cascade that starts at stock A, which is denoted by the large blue star?' and ii) 'If so, how exactly will stock Z be affected (buy vs. sell)?'

On one hand, you can immediately *see* how easy it is to answer the first question. The middle panels show that there's a path connecting stock A to stock Z in M_2, M_3, and M_4 but not in M_1. So, stock Z might be affected by an index-fund rebalancing cascade starting with stock A in M_2, M_3, and M_4 but not in M_1. Answering this first question gives you a sense of what it means to have a polynomial-time solution.

But, on the other hand, you can also immediately *see* how hard it is to answer the second question. There's no way to guess how an index-fund rebalancing cascade will affect stock Z by examining the set of index-fund rebalancing rules involved, even though these rules are completely deterministic. M_2, M_3, and M_4 all have paths connecting stock A to stock Z ending positive shocks. But, the effect of the entire index-fund rebalancing cascade only agrees with this naïve prediction in M_2.

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

Problem 2.3a (How).

- *Instance: A choice for stock Z; a market structure M; a time $T = \mathrm{Poly}[S]$; a positive constant $\epsilon > 0$; and, the power set $\hat{\mathsf{A}} \subseteq 2^{\mathsf{S}}$ of all ϵ-small sets $\mathsf{A} \subseteq \mathsf{S}$.*

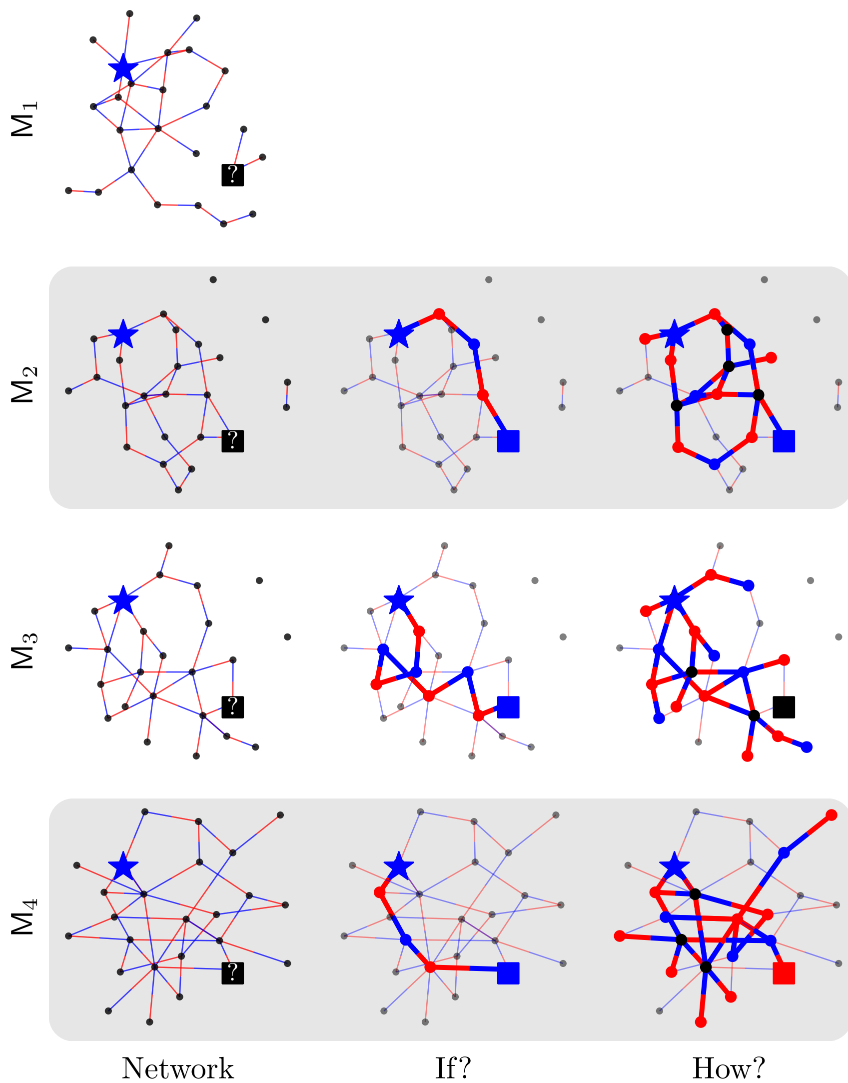


Figure 3: Some Intuition. *Each row contains 3 panels and depicts simulated results for a single market with* $S = 25$ *stocks—i.e., one market structure per row. Nodes are stocks. Node color denotes effect of index-fund rebalancing cascade: blue=positive, red=negative, black=no effect. Star: stock A.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

- *Question: Is there some* $\mathsf{A} \in \hat{\mathsf{A}}$ *such that* $\text{Effect}_{\mathsf{M},T}(\mathsf{A}, Z) \neq +1$*?*

How denotes the set of instances where the answer is 'Yes'. Here's what How is asking in plain English. Imagine the universe of all index-fund rebalancing cascades that stem from an initial positive shock to an arbitrarily small subset of stocks in the market. Will every single one of these rebalancing cascades have a positive effect on stock Z after T rounds of rebalancing?

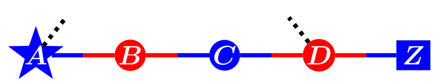
How *Complexity.* The proposition below gives a mathematical result that mirrors the intuition we built up in Figure 3. Solving How is much harder than solving If.

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

Just like instances of If, the size of instances of How are governed by the number of stocks 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's hard to solve but easy to verify (Garey and Johnson, 2002). Solving this Sunday's grid might take you an hour, but it will only take a second to verify your guess for 31-down using the answer key.

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 every decision problem in NP can be reduced to it and it belongs to NP.

Root of the Problem. Figure 4 illustrates why How is so computationally intractable. Each vertical gray region denotes a separate sequence events, starting at the top and ending at the bottom. On the left, there's a proposed path connecting stock A to stock Z that ends with a positive shock to stock Z:



The trouble is that stocks A and D are also connected to other stocks that may not belong to the original path (dotted lines), which 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 cascade's net effect on stock Z. If stock A and stock D are directly connected, M_0, then the secondary path doesn't matter. If there is a 1-path connecting stock A to stock D, M_1, then

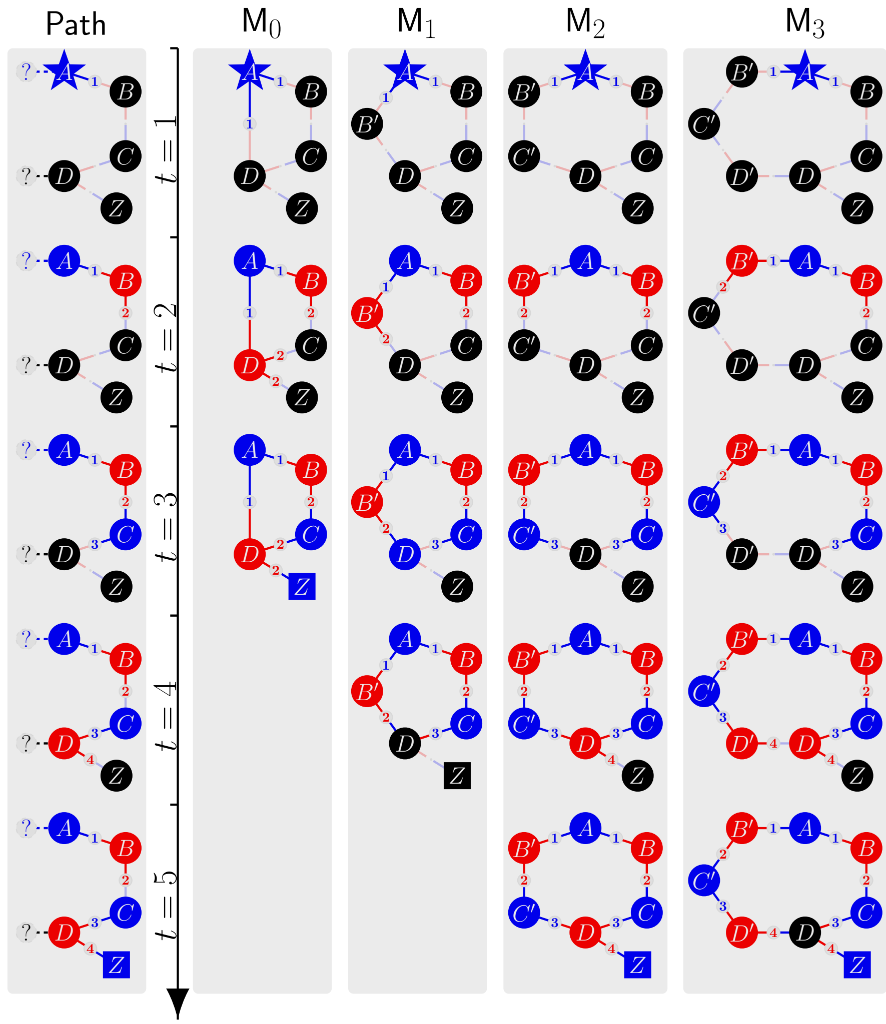


Figure 4: Root of the Problem. *Each vertical gray region denotes a separate sequence events, which starts at the top and ends at the bottom. Each node denotes a stock. Node color denotes effect of cascade: blue=positive, red=negative, black=no effect. Star: initial shock to stock A. Square: final effect for stock Z. Edges denote index-fund rebalancing rules. Blue(s)-to-red(s'): stock s' is negative neighbor to stock s. Red(s)-to-blue(s'): stock s' is positive neighbor to stock s.* Path*: path connecting stock A to stock Z. Location of stocks A, B, C, D, and Z remain unchanged in all sequences. Dotted lines: neighbors to stock A and stock Z that could form alternate path.* M_k*: market structure that contains alternate path with $k \in \{0, 1, 2, 3\}$ stocks separating stock A and stock D.*

the secondary path implies that stock Z will be unaffected by the entire index-fund rebalancing cascade. But, if there's a 2-path connecting stock A to stock D, M_2, then the secondary path won't matter once again. And, if there's a 3-path connecting stock A to stock D, M_3, then stock Z will be positively affected by the index-fund rebalancing cascade even though stock D will be unaffected. Tiny changes in the structure of a rebalancing cascade can lead to different outcomes for stock Z.

As a result, determining how a particular index-fund rebalancing cascade will affect stock Z requires a detailed simulation of how the entire cascade will play out. So, finding an initial shock which results in a negative effect on stock Z could require checking every possible ϵ-small subset. And, the size of this power set scales exponentially with the number of stocks in the market, S. Suppose you could solve instances of How in less than one microsecond when there were only 20 ETFs in the market. Proposition 2.3a implies that this same algorithm would take longer than the current age of the universe to execute in today's market, which contains roughly 2,000 U.S.-listed ETFs.[6] "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)."

Predicting How. Proposition 2.3a says that the problem of figuring out how every single index-fund rebalancing cascade will effect stock Z is computationally intractable. But, maybe this is an unreasonable goal. What if you only try to figure out how most index-fund rebalancing cascades will affect stock Z? We introduce the following decision problem to make this idea precise.

Problem 2.3b (MajorityHow)**.**

- *Instance: A choice for stock Z; a market structure M; a time $T = \mathrm{Poly}[S]$; a positive constant $\epsilon > 0$; and, the 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, MajorityHow seems like a much closer analogue to the problem that real-world traders care about. Traders know which index funds hold each stock. And, they know the rebalancing rules that index funds are following. So, given this information, they would like to determine whether or not some stock Z will be affected by the majority of index-fund rebalancing cascade that might occur. For a particular market structure, will more than half of all possible index-fund rebalancing cascades

[6]Financial Times. 1/14/2017. *ETFs Are Eating The US Stock Market.*

result in buy orders?

At first, MajorityHow seems like a much easier problem to solve than How because it doesn't involve finding a particular verboten instance. But, this first reaction is wrong. Proposition 2.2b shows that stock Zs with more neighbors are more likely to be hit by index-fund rebalancing cascades. But, Proposition 2.3b shows that determining whether more than half of all possible index-fund rebalancing cascades will result in buy orders 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. And, a polynomial-time solution to MajorityHow would imply $\mathsf{P} = \mathsf{NP}$.

2.4 Key Ingredients

We've just seen that predicting how index-fund rebalancing cascades will affect a stock's demand with accuracy better than a coin flip is an NP-hard problem. As a result, the demand shocks coming from the rebalancing cascades are effectively noise. To make it easier for other researchers to spot other situations where the same mathematical reasoning applies, we now describe three key features of index-fund rebalancing cascades that make them so hard to predict.

Alternation. First, index-fund rebalancing cascades are only hard to predict if they involve alternating sequences of buy and sell orders. In a world where a positive shock to stock A can only ever result in a positive shock to stock B, predicting how stock Z will be affected by a rebalancing cascade is easy. In fact, it's equivalent to solving the 'If?' problem.

Proposition 2.4a (Necessity of Alternation)**.** *Without alternation,* How *is solvable in polynomial time.*

Index-fund rebalancing cascades necessarily involve an alternating sequence of buy and sell orders. When an index fund rebalances its portfolio, it swaps out an existing position in one stock for a new position in another. But, there are other cascade-like phenomena where this isn't the case. For example, think about bank runs. During a bank run, depositors are choosing whether to withdraw their money—sell only. As a result, equilibrium demand in these models behaves in a predictable way depending on whether some threshold has been crossed (Diamond and Dybvig, 1983).

Feedback Loops. Second, index-fund rebalancing cascades are only hard to predict in a market structure that involves cancellation due to feedback loops. It's important that different cascade paths have the potential to cancel each other out, as shown in Figure 4. To illustrate, think about what would happen if every stock in the market had exactly 2 neighbors. In this setting, if there exists a path connecting stock A to stock Z, then you can determine how a rebalancing cascade starting with stock A will affect stock Z by counting the number of stocks in the path. If it's an odd number, then stock Z will realize a positive shock, like in Equation (11). Whereas, if it's an even number, then the shock will be negative, like in Equation (12).

Proposition 2.4b (Necessity of Feedback Loops)**.** *Without cancellation due to feedback loops,* How *is solvable in polynomial time.*

Again, we feel that feedback loops are a natural part of the index-fund universe. There is no central-planning committee that limits the number of indexes that a single stock can belong to. There's nothing stopping 20 different smart-beta ETFs from holding the same stock at the same time.[7] Thus, the associated collections of rebalancing rules will contain market structures with feedback loops. And, it's these loopy instances that make solving How computationally intractable.

Thresholds. Third, index-fund rebalancing cascades are only hard to predict if their benchmark indexes involve threshold-based rebalancing rules. For example, it's 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 and not a benchmark including all S&P 500 stocks with relatively more shares of lower-volatility constituents. In the first case, an arbitrarily small change in a stock's volatility can move it from 101st to 100th place on the low-volatility leaderboard and force SPLV to exit its entire position. In the second case, an arbitrarily small change in a stock's volatility would only lead to an even smaller change in the fund's portfolio position. Without threshold-based rebalancing rules, longer cascade paths would necessarily have smaller effects for the same reason that AR(1) impulse-response functions get weaker at longer horizons. So, you could approximate an index-fund rebalancing cascade's net effect on stock Z by using the effect of the shortest path to stock Z.

Proposition 2.4c (Necessity of Thresholds)**.** *If index funds don't use threshold-based rebalancing rules, then there's a fully polynomial-time approximation scheme for* How*.*

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

It's a simple fact that index funds use threshold-based rebalancing rules. This is how many index funds operate. But, threshold-based trading rules can be found all over the place in financial markets. 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. Our goal is not to explain why funds choose to follow these sorts of trading rules; instead, we point out one natural consequence of this choice: noise.

No-Trade Theorem. We began this paper with a discussion of Milgrom and Stokey (1982)'s classic no-trade theorem. There's no error in their paper. So, at this point, you might be wondering why doesn't their result apply to the setting we study in our paper. What implicit assumption is being violated?

Milgrom and Stokey (1982) consider a setting where all traders start out with common priors and then one of them gets a private signal. They then prove that, if this lone trader acts on his private signal using a simple deterministic trading rule, then everyone else in the market will be able to figure out what he's learned by studying his trading behavior. We show that this result can break down in modern financial markets because there isn't just one lone trader following a simple deterministic trading rule. There are hordes of them. So, even if each index fund is using a simple deterministic rebalancing rule, the net demand coming from the entire interacting mass of index funds can still appear random.

Different Application. Finally, we would like to point out a nice parallel between our main theoretical results and the analysis in Arora et al. (2011). Instead of studying the demand-shock distribution for a single stock, Arora et al. study the loan-quality distribution within a single mortgage-backed security. They too show that the problem of determining whether an asset-backed security contains slightly more bad loans than expected is NP hard. Same mathematical insight. Different financial applications.

3 Rebalancing Cascades

We've just seen that index-fund rebalancing cascades can generate seemingly random demand shocks in a theoretical model. We now use data on end-of-day ETF holdings to show that the ETF rebalancing cascades generate unpredictable demand shocks in real-world financial markets.

3.1 Index Funds

We study index-fund rebalancing cascades using data on a particular kind of index fund—namely, exchange-traded funds (ETFs). There are three reasons for this choice.

Reason #1: Diversity. First, we need a large group of index funds that follows a very heterogeneous collection of benchmark indexes. Prior to January 2008, ETFs all looked like the SPDR S&P 500 ETF [`SPY`] in that they all tracked some sort of pre-existing market index, like the S&P 500. But, in early 2008, the SEC changed its guidelines so that an ETF could track its own self-defined benchmark. After this change, Invesco PowerShares was free to create an ETF tracking the returns of the quintile of S&P 500 stocks with the lowest historical volatility even though there was no pre-existing low-volatility S&P 500 index. All Invesco had to do was promise to announce the identities and weights involved in the benchmark one day in advance.

Now, there are more ETFs than stocks.[8] "From ProShares we have `CLIX` (100% long internet retailers and 50% short bricks-and-mortar U.S. retailers) and `EMTY` (which just bets against bricks-and-mortar retailers)... meanwhile from EventShares, we have policy-factor ETFs... like... `GOP` (full of oil drillers, gun manufacturers, and so on that would benefit from Republican policies) and `DEMS` (with companies that should do well under Democrats, such as clean-energy companies). There is also an ETF called `TAXR` that invests in companies poised to benefit most from a successful attempt to pass a tax reform bill."[9]

The sheer number and variety of these so-called 'smart-beta' ETFs has become something of a hot-button issue of late. To be sure, niche ETFs like `DEMS` tend to be smaller than broad value-weighted market ETFs, like the SPDR S&P 500 ETF `SPY`. But, even the rebalancing activity of niche ETFs can affect a stock's fundamentals because ETFs often execute the bulk of their trades during the final 20 to 30 minutes of the trading day. 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."[10]

Reason #2: Discretion. Second, ETF managers have less ability to deviate from their stated benchmarks than either mutual- or hedge-fund managers due to the

[8]Bloomberg. 5/16/2017. *Mutual Funds Ate the Stock Market. Now ETFs Are Doing It.*
[9]Financial Times. 11/21/2017. *A ROSE by any other ticker symbol...*
[10]Financial Times. 3/17/2017. *Machines and Markets: 5 Areas To Watch.*

underlying structure of the ETF market (Madhavan, 2016; Ben-David et al., 2017). The company running an ETF (its 'sponsor') has an obligation to create or redeem shares at the end-of-day market value of its stated benchmark. So, if an ETF's price is higher than the end-of-day market value of its benchmark, then an arbitrageur can sell shares of the ETF back to the sponsor and use the proceeds to buy shares of the underlying assets in the benchmark index. The reverse logic holds when underpriced.

If arbitrageurs are constantly asking an ETF sponsor to create or redeem lots of shares, then the sponsor must be losing lots of money. So, just like you'd expect, creations and redemptions are only a small fraction of daily trading volume for ETFs, and these trades involve less than 0.5% percent of ETFs' net assets (Investment Company Institute, 2015). Instead, ETF trading volume primarily comes from managers' rebalancing activity just prior to market close. This end-of-day trading is how ETF sponsors make sure that there is very little difference between the market value of their end-of-day holdings and the market value of their stated benchmark.

An ETF manager who does the bulk of his rebalancing right at market close will incur higher trading costs. But, the typical investor in a smart-beta ETF is not looking for a cheap way to buy and hold a broad market portfolio. ETF investors traded "$20 trillion worth of shares last year even though ETFs only have $2.5 trillion in assets. That's 800% asset turnover, which is about 3-times more than stocks."[11] An investor interested in holding a smart-beta ETF is looking for quick access to a very targeted position. He'd rather the ETF manager have slightly higher trading costs and be much more faithful to his stated benchmark. For a niche ETF, the additional trading costs incurred by the end-of-day trading are nothing compared to the costs associated with replicating the entire position from scratch.

Reason #3: Data. Third, we can observe end-of-day portfolio positions for ETFs. Specifically, we use data from ETF Global that includes both the assets under management, $AUM_{f,t}$, and the relative portfolio weight on each stock, $\Omega_{f,s,t}$, for each ETF $f \in \{1, \ldots, F\}$ at the end of each trading day from January 2010 to December 2015. We restrict our sample to include only those ETFs that rebalance their positions daily—think about the PowerShares S&P 500 Low-Volatility ETF [`SPLV`] rather than the SPDR S&P 500 ETF [`SPY`]. 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, $Q_{f,s,t}$, is:

$$Q_{f,s,t} = \tfrac{1}{P_{s,t}} \times \left\{ \Omega_{f,s,t} \cdot AUM_{f,t} \right\} \tag{15}$$

[11]Bloomberg. 3/3/2017. *5 Ways 'Passive' Investing Is Actually Quite Active.*

And, total ETF trading volume for stock s on day t is given by $\sum_{f=1}^{F} |\, Q_{f,s,t} - Q_{f,s,t-1} \,|$.

This end-of-day data is important. Other papers in the ETF literature, such as Ben-David et al. (2017), impute ETFs' daily portfolio positions from their end-of-quarter financial statements. But, we are interested in how the rebalancing decisions of different ETFs interact with one another over the course of a few days. And, we can't study these interactions if we are forced to impute daily holdings from end-of-quarter data.

Rebalancing Volume. We use data on end-of-day ETF holdings to create two main variables of interest. The first is ETF rebalancing volume. This requires a little bit of subtlety because not all ETF trading is due to rebalancing decisions. If money pours into ETF f on day t, $AUM_{f,t} \gg AUM_{f,t-1}$, then the fund is going to have to buy a bunch of shares of each stock that it holds. But, this trading volume won't be due to any rebalancing decision. The fund manager is just scaling up his existing holdings. So, to adjust for trading due to inflows and outflows, we first calculate each ETF's predicted holdings on day t given its portfolio weights on the previous day $(t-1)$ and the realized inflows and outflows on day t:

$$\bar{Q}_{f,s,t} = \tfrac{1}{P_{s,t}} \times \left\{ \Omega_{f,s,t-1} \cdot AUM_{f,t} \right\} \tag{16}$$

Then, for each stock s, we compute the total difference between every ETF's actual end-of-day holdings and this inflow-adjusted prediction on day t:

$$\texttt{etfRebalVlm}_{s,t} = \textstyle\sum_{f=1}^{F} \left|\, Q_{f,s,t} - \bar{Q}_{f,s,t} \,\right| \tag{17}$$

We use this as our daily measure of ETF rebalancing volume for each stock. Note that we will write all regression variables in `teletype font` to distinguish them from estimated parameters. Table 2 shows that the typical stock during our sample period had $e^{7.62} \approx 2{,}038$ shares traded each day due to ETF rebalancing decisions.

Order Imbalance. Then, to evaluate whether ETFs are trading in different directions, we also compute a corresponding measure of ETF order imbalance:

$$\texttt{etfOrdImbal}_{s,t} = \sum_{f=1}^{F} \frac{Q_{f,s,t} - \bar{Q}_{f,s,t}}{\texttt{etfRebal}_{s,t}} \tag{18}$$

This variable lies on the interval $[-1, 1]$. If $\texttt{etfOrdImbal}_{s,t} = -1$, then every share of stock s traded on day t was a sell order. Whereas, if $\texttt{etfOrdImbal}_{s,t} = 1$, then every share of stock s traded on day t was a buy order. Table 2 contains summary statistics describing the typical ETF order imbalance for each stock during our sample period.

3.2 Initial Shocks

We use M&A announcements for our set of initial shocks, referring to the stock that's the target of the M&A announcement as stock A.

M&A Announcements. Our source for data on M&A deals is Thomson Financial. We use all deals with an announcement date between January 1st, 2010 and December 31st, 2015 where the target is a public company. Table 2 shows that there were 884 such events during our sample period. M&A announcements are a natural choice for our initial shocks because there is solid empirical evidence that the target of an M&A announcement realizes a sharp price increase (Andrade et al., 2001). And, while acquirers do not choose their M&A targets at random, the exact day that a deal is announced can be taken as random. We use t_A to denote the day of the M&A announcement in which stock A was the target firm.

Desired Effect. Table 3 contains direct evidence that ETFs rebalance in response to these initial shocks. We create a panel dataset containing the ETF rebalancing volume for each M&A target in our sample during the time window $t \in \{t_A - 20, \ldots, t_A + 5\}$. Then, we regress log ETF rebalancing volume on indicator variables for the date of the M&A announcement:

$$\ln(\texttt{etfRebal}_{A,t}) = \alpha + \beta \cdot 1_{\{t=t_A-1\}} + \gamma \cdot 1_{\{t=t_A\}} + \delta \cdot 1_{\{t=t_A+1\}} + \cdots + \varepsilon_{A,t} \quad (19)$$

In different specifications, the "$\cdots$" term in the equation above contains year-month fixed effects, stock A fixed effects, and lagged trading volume. The first column in Table 3 shows that ETF rebalancing volume for stock A rises by 166% on the day it's announced as an M&A target. The second column in Table 3 shows that this jump in ETF rebalancing activity is not explained by lagged volume.

Manager Discretion. The third column of Table 3 shows the results of the same regression specification but with additional indicator variables for days $(t_A - 2)$, $(t_A - 3)$, $(t_A - 4)$, and $(t_A - 5)$. This column reveals that there is no pretrend in ETF managers' reaction to the M&A announcement. ETF rebalancing volume only starts to rise on the day immediately before the announcement, and this 1-day-early effect is due to the way overnight announcements are coded by Thompson Financial. What's more, after we include lagged volume in the specification, the jump in ETF rebalancing volume is gone the day after. This supports our claim that ETF managers don't have much discretion when it comes to deviating from their benchmark index overnight. There's no reason to suspect that ETF managers are slowly rebalancing their position in stock Z in response to demand shocks coming from ETF rebalancing cascades if they aren't doing the same thing in response to the initial M&A announcement shock to stock A.

Placebo Test. Finally, the fifth column of Table 3 shows the results of the same regression specification using randomly selected dates for t_A rather than the actual M&A announcements dates. Just as expected, there is no sharp jump in ETF rebalancing volume on these randomly selected dates. But, more importantly, the coefficients on lagged trading volume are also unchanged. This placebo test suggests that our result isn't being driven by some broader trading-volume anomaly that occurs around the time of each M&A announcement. Things look normal other than the spike in ETF rebalancing volume.

3.3 Main Results

We now give evidence that these M&A announcements lead to ETF rebalancing cascades that result in unpredictable demand shocks for unrelated stock Zs.

Diff-in-Diff Approach. Here's how we set up our tests. First, we create a panel dataset containing the ETF rebalancing volume on days $t \in \{t_A - 20, \ldots, t_A + 5\}$ for each unrelated stock Zs relative to each M&A target announcement in our sample. We use $\texttt{afterAncmt}_{A,t}$ as an indicator variable to flag the 5 days after the M&A announcement for a given stock A:

$$\texttt{afterAncmt}_{A,t} = \begin{cases} 1 & \text{if } t \in \{t_A + 1, \ldots, t_A + 5\} \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

For stock Z to be unrelated to stock A, it has to be twice removed in the network of ETF holdings. It can't have been recently held by any ETF that also recently held stock A. And, if stock B and stock A are both held by the same ETF, then stock Z can't have been recently held by any ETF that also recently held stock B either. i.e., the chain has to be $A \to B \to C \to Z$ or longer. Because there are so many different smart-beta ETFs that are specifically designed to give their investors exposure to things like size and value, this criteria implies that each set of stock Zs doesn't share well-known characteristics with the associated stock A.

Proposition 2.2b suggests that, all else equal, stocks on the cusp of more rebalancing thresholds are more likely to be hit by an ETF rebalancing cascade. So, we split the set of stock Zs for each initial M&A announcement into two subsets: those on the cusp of an above-median number of ETF rebalancing thresholds (i.e., stocks with lots of neighboring stocks in the ETF rebalancing network) and those on the cusp of

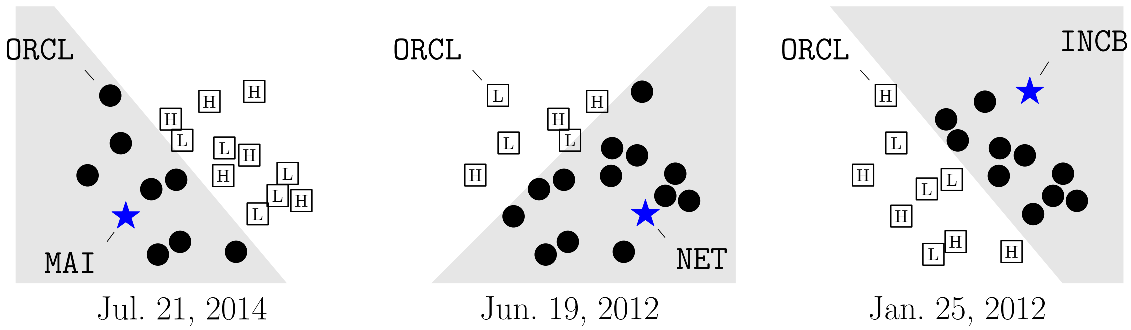


Figure 5: Empirical Design. *Each panel depicts the the same set of stocks during 3 different M&A announcements: Owens & Minor's purchase of Medical Action Industries* [`MAI`] *announced on Jul. 21, 2014; Sonus Networks' purchase of Network Equip Technologies* [`NET`] *announced on Jun. 19, 2012; and, 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's related to stock A at the time of the announcement. Each white square denotes a stock that's 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.*

a below-median number of ETF rebalancing thresholds. We use $\texttt{manyNbrs}_{A\to Z,t}$ as an indicator variable to flag the subset of stock Zs that are on an above-median number of ETF rebalancing thresholds:

$$\texttt{manyNbrs}_{A\to Z,t} = \begin{cases} 1 & \text{if stock } Z \text{ has an above-median number of neighbors} \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

We say that stock s' is a neighbor to stock s if a fund that currently holds stock s also rebalanced its position in stock s' at some point during the previous month.

There are two key predictions from Section 2 that we want to test. First, Proposition 2.2b suggests that stock Zs with more neighbors should be more likely to be hit by ETF rebalancing cascades and so should have proportionally higher ETF rebalancing volume in the days immediately following the initial M&A announcement for stock A. We test this prediction using a standard diff-in-diff regression:

$$\begin{aligned} \ln(\texttt{etfRebalVlm}_{Z,t}) = \alpha + \beta \cdot \texttt{afterAncmt}_{A,t} & \\ + \gamma \cdot \texttt{manyNbrs}_{A\to Z,t} & \\ + \delta \cdot \{\texttt{afterAncmt}_{A,t} \times \texttt{manyNbrs}_{A\to Z,t}\} & \\ + \cdots + \varepsilon_{A\to Z,t} & \end{aligned} \tag{22}$$

The null hypothesis is that ETF rebalancing cascades only affect stocks that are closely related to the initial stock A. If this were the case, then we'd expect to find $\delta = 0$. By contrast, if stock Zs with more neighbors are indeed more likely to be hit by a rebalancing cascade, then we should estimate $\delta > 0$.

How should we interpret β? ETF rebalancing cascades have the potential to affect the demand for all stock Zs; it's just that they're much more likely to affect the demand of stock Zs with more neighbors. Stocks with only 1 or 2 neighbors can still be included in a rebalancing cascade as shown in Figure 4. So, if ETF rebalancing cascades are taking place, then we should also expect to find $\beta > 0$. And, we can use this auxiliary prediction as a way of checking the internal consistency of our results. But, while finding $\beta > 0$ is consistent with the existence of ETF rebalancing cascades, it's also consistent with general market conditions changing following an M&A announcement. By checking whether $\delta > 0$, we can see whether this increase in ETF rebalancing volume for stock Z is related to the structure of ETF rebalancing rules in a way that's predicted by our theory.

Second, even though stock Zs with more neighbors are more likely to be hit by ETF rebalancing cascades, Proposition 2.3b suggests that the direction of the resulting demand shock should be a coin flip. We test this prediction using the same diff-in-diff specification as before but with order imbalance as the left-hand-side variable:

$$\begin{aligned}\ln(\texttt{etfOrdImbal}_{Z,t}) = \alpha + \beta \cdot \texttt{afterAncmt}_{A,t} \\ + \gamma \cdot \texttt{manyNbrs}_{A \to Z,t} \\ + \delta \cdot \{\texttt{afterAncmt}_{A,t} \times \texttt{manyNbrs}_{A \to Z,t}\} \\ + \cdots + \varepsilon_{A \to Z,t}\end{aligned} \tag{23}$$

If a stock Z with many neighbors is no more likely to realize a positive demand shock than a stock Z with few neighbors, then we should estimate $\delta = 0$.

Empirical Design. At this point, you might be worried that the unrelated stock Zs with more neighbors are just different kinds of stocks than the unrelated stock Zs with few neighbors. And, this is a valid concern. But, there is an important detail about how we set up our diff-in-diff approach that helps us address this concern. Specifically, we define the set of unrelated stock Zs separately for each initial M&A announcement for a stock A. This means that the exact same stock can play the role of an above-median stock Z relative to one M&A announcement while playing the role of a below-median stock Z relative to another. Figure 5 gives an example from

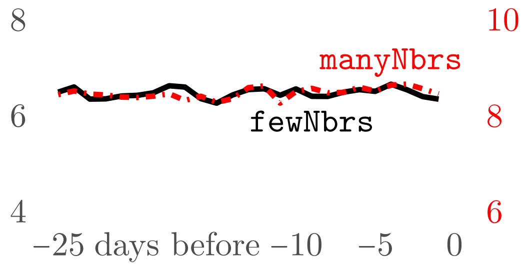


Figure 6: No Pre-Trend. *x-axis: days before stock A is announced as M&A target. y-axis: log ETF rebalancing volume each day for stock Zs that are unrelated to stock A. Solid-black line (left axis) corresponds to subset of stock Zs with few neighbors. Dashed-red line (right axis) corresponds to subset with many neighbors.*

our dataset of this sort of thing happening for Oracle Corp.

By including fixed effects for each stock Z in our regression specification, we can estimate how the ETF rebalancing activity for the exact same stock changes when it happens to have many neighbors. This design feature means that any empirical results we find can't be explained by ETFs always trading some stocks differently than others. Any confounding variable has to explain why ETFs suddenly change their rebalancing behavior for an ever-changing subset of stock Zs in the days immediately following an M&A announcement about a completely unrelated stock A.

Rebalancing Volume. Table 4 provides the estimated coefficients for the regression described in Equation (22). The first column shows that ETF rebalancing volume for all unrelated stock Zs tends to rise by 5.0% on average in the wake of an M&A announcement for stock A. But, the third column shows that this growth is concentrated among unrelated stock Zs that have many neighboring stocks in the network defined by ETF rebalancing rules. Consistent with our economic story, we find that ETF rebalancing volume is 3.7% higher for the above-median group of stock Zs than for the below-median group in the five days immediately following each M&A announcement.

There are three important points to emphasize about this result. The first is that we include stock-Z fixed effects in our regressions. So, because the same stock Z can have many neighbors relative to one M&A target and few neighbors relative to another, these results can't be due to persistent differences in how ETFs tend to rebalance their positions in particular stocks. The second is that there is no pre-trend. Figure 6 shows that in the run-up to each M&A announcement, the difference between the amount of ETF rebalancing activity in stock Zs with many neighbors and the amount of ETF rebalancing activity in stock Zs with few neighbors remains constant. Finally, the second and fourth columns of Table 4 confirm that the sudden spike in log ETF rebalancing volume for stock Zs with many neighbors isn't due to a general run-up in trading volume. When we include lagged trading volume in our regression specification, our point estimates are largely unchanged.

Order Imbalance. Table 4 gives evidence of long rebalancing cascades; whereas, Table 5 gives evidence that these cascades have an unpredictable effect on demand. This table reports the estimated coefficients for the regression described in Equation (23). The main takeaway from this table comes from comparing the coefficients in the first and third columns. While the first column shows that there is a statistically significant $\beta = 0.0075\%$pt increase in ETF order imbalance for unrelated stock Zs in the days immediately after an M&A announcement, the third column shows that there is no measurable difference between the ETF order imbalance of stock Zs with many neighbors and stock Zs with few neighbors. What's more, the size of the statistically insignificant point estimate for this difference, $\delta_{\texttt{etfOrdImbal}} = 0.0024\%$pt, is more than 2 orders-of-magnitude smaller than the corresponding difference in ETF rebalancing volume, $\delta_{\ln(\texttt{etfRebal})} = 3.70\%$. Taken together, this evidence suggests that, while it's possible to predict which stock Zs are likely to be affected by a ETF rebalancing cascade, it's much harder to predict how these stock Zs will be affected by the resulting demand shock.

Aggregate Tests. Among empirical economists, it's taken almost as an article of faith that empirical tests should be run using the most micro-level data possible. So, at this point, you might be surprised that we didn't try to trace out the precise buy-sell-buy-sell sequences of each ETF cascade in our sample. But, there is a good reason why we didn't do this. This empirical approach would fundamentally ignore the central message of our theoretical analysis: it is computationally intractable to make predictions about the fine-grained structure of rebalancing cascades. Instead, we need to run our tests using well-chosen macro-level variables. Even if it isn't practical to track the precise buy-sell-buy-sell sequence of ETF rebalancings, it's relatively easy to proxy for the total number of thresholds that a stock is close to. By analogy, even if it isn't possible to keep track of the location and momentum of every single gas molecule in a 1m^3 box, it's easy to measure macro-level variables like the pressure and temperature inside the container.

4 Market Reaction

We've just seen evidence that ETF rebalancing cascades exist and that the resulting demand shocks are unpredictable from the standpoint of an econometrician. But, maybe these demand shocks look less random to traders? In this section, we provide evidence that market participants treat the erratic demand coming from ETF rebalancing cascades as noise by studying cross-sectional variation in liquidity.

4.1 Liquidity

If market participants treat the erratic demand coming from ETF rebalancing cascades as noise, then we should find that stock Zs with more neighbors also have higher levels of liquidity.

Variable Definitions. We calculate the liquidity of each stock Z on a daily basis in two different ways. First, we compute an intraday variant of the Amihud (2002) measure:

$$\texttt{amihud}_{Z,t} = \frac{1}{390} \cdot \sum_{m \in t} \frac{|R_{Z,m}|}{\$\mathrm{V}_{Z,m}} \tag{24}$$

Above, $m \in \{1, \ldots, 390\}$ indexes the 390 minutes in each trading day, $\mathrm{ret}_{Z,m}$ denotes the return of stock Z in minute m, and $\$\mathrm{V}_{Z,m}$ denotes the number of dollars of stock Z traded in minute m. We scale this variable so that it's reported as the percent-change in stock Z's price per million dollars of traded volume. In addition, we also compute the average bid-ask spread of stock Z during the trading day:

$$\texttt{baSpread}_{Z,t} = \frac{1}{P_{Z,t}} \cdot \left(\frac{1}{390} \cdot \sum_{m \in t} [P^{\text{bid}}_{Z,m} - P^{\text{ask}}_{Z,m}] \right) \tag{25}$$

We scale this second variable so that it's reported in basis points as a fraction of stock Z's closing price on day t.

Around Initial Shocks. Market participants can see whether each unrelated stock Z has many neighbors or few neighbors. And, because the stock Zs with many neighbors are more likely to be hit by an ETF rebalancing cascade, market participants should realize that they are more likely to see erratic non-fundamental demand shocks for these stocks as a result. So, the stock Zs with many neighbors should have higher liquidity. The second and fourth columns of Table 6 confirm this prediction by showing that $\gamma < 0$ when estimating the regression below for $\texttt{y} \in \{\texttt{amihud}, \texttt{baSpread}\}$:

$$\begin{aligned} \texttt{y}_{Z,t} = \alpha &+ \beta \cdot \texttt{afterAncmt}_{A,t} \\ &+ \gamma \cdot \texttt{manyNbrs}_{A \to Z,t} \\ &+ \delta \cdot \{\texttt{manyNbrs}_{A \to Z,t} \times \texttt{afterAncmt}_{A,t}\} \\ &+ \cdots + \varepsilon_{A \to Z,t} \end{aligned} \tag{26}$$

Note that both the Amihud (2002) measure and the bid-ask spread are inversely related to liquidity. So, $\gamma < 0$ implies that stock Zs who happen to have more neighbors also have more liquidity.

These same columns also show that there is no change in the relative liquidity of the stock Zs with many neighbors and the stock Zs with few neighbors in the days immediately after the initial M&A announcement for stock A. This result is in stark contrast to the earlier findings in Tables 4 and 5. But, this result is also exactly what we'd expect to find in a market where traders knew which stock Zs were most susceptible to the erratic non-fundamental demand coming from ETF rebalancing cascades. This knowledge should be priced into each stock Z's bid-ask spread ahead of time. The market as a whole might contain more asymmetric information after an important M&A announcement and thus be less liquid in general. But, if market participants already understood which stocks were more likely to be hit by ETF rebalancing cascades and considered this demand to be noise, then there should be no differential effect for stock Zs with many neighbors vs. those with few neighbors following the M&A announcement for stock A.

4.2 Panel Regressions

If market participants recognize ex ante that stocks with more neighbors are more likely to be hit by ETF rebalancing cascades, then you might expect that stocks with many neighbors actually have higher liquidity than stocks with few neighbors unconditionally. And, this is exactly what we find in the data.

Unconditional Results. The third and fifth columns of Table 7 show the estimation results for the regression below where $\mathtt{y} \in \{\mathtt{amihud}, \mathtt{baSpread}\}$:

$$\mathtt{y}_{s,t} = \alpha + \beta \cdot \mathtt{\#nbrs}_{s,t} + \cdots + \varepsilon_{s,t} \tag{27}$$

The key difference between this regression and the earlier regressions is that the data we use to estimate this regression do not only include the unrelated stock Zs around each initial M&A announcement. They include all stocks in our data sample. Again, we estimate $\beta < 0$ for both liquidity measures, suggesting that market participants are treating the erratic demand coming from ETF rebalancing cascades as noise.

Implication for Traders. A natural next question is: 'What should a trader do with this information?' The answer isn't to directly buy or sell stocks with many or few neighbors. Instead, these results suggest a way of amplifying the returns to any existing cross-sectional trading strategy. For example, suppose that you would like to construct a classic momentum portfolio that is long the 30% of stocks with the highest returns over the previous 6 months and short the 30% of stocks with the lowest returns over the previous 6 months. Our results suggest that you could implement this strategy more efficiently by focusing each leg of this strategy on the stocks with the most neighbors. This position will have the same average return, but it will have lower implementation costs due to the liquidity provided by ETF rebalancing cascades.

Neighbors vs. Holdings. Are our results just due to the fact that ETFs like holding liquid stocks? No. And, we can use the regression specification in Equation (27) one more time to further emphasize this point. Our results are due to the overlapping network of ETF rebalancing decisions and not due to which stocks ETFs choose to hold. Specifically, we re-estimate the regression where $\mathtt{y} \in \{\mathtt{amihud}, \mathtt{baSpread}\}$, but this time including fixed effects for the number of ETFs that hold a particular stock:

$$\mathtt{y}_{s,t} = \alpha + \beta \cdot \mathtt{\#nbrs}_{s,t} + \gamma \cdot 1_{\{\mathtt{\#etf}=i\}} + \cdots + \varepsilon_{s,t} \tag{28}$$

By doing this, we are able to estimate the amount of additional liquidity that is associated with having more neighbors while controlling for the effect of being held by more ETFs. Again, we find that $\beta < 0$, suggesting that market participants are reacting to a stock's susceptibility to ETF rebalancing cascades rather than just the number of ETFs that hold it.

5 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 seemingly random demand shocks in financial markets. A stock's demand might appear random, not because individual investors are behaving randomly, but because it's too computationally complex to predict how a wide variety of simple, deterministic, trading rules will interact with one another. First, we show theoretically how computational complexity can generate noise by modeling a particular kind of trading-rule interaction: index-fund rebalancing cascades. Then, we give empirical evidence that index-fund rebalancing cascades actually generate noise in real-world financial markets using data on the end-of-day holdings of exchange-traded funds (ETFs).

By showing precisely why it's computationally intractable to predict ETF rebalancing cascades, we make it possible for researchers to identify other situations where the same logic holds. For example, our theoretical model also applies to any other group of funds following a wide variety of threshold-based rebalancing rules. Think about quantitative hedge funds following strategies of the form 'Buy the top 30% and sell the bottom 30% of stocks when sorting on X' (Khandani and Lo, 2007). Or, consider pension funds with strict portfolio mandates of the form '15% of our assets will be held in alternative investments' (Pennacchi and Rastad, 2011).

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*.

Arora, S., B. Barak, M. Brunnermeier, and R. Ge (2011). Computational complexity and information asymmetry in financial products. *Communications of the ACM*.

Atkeson, A., A. Eisfeldt, and P. Weill (2015). Entry and exit in OTC derivatives markets. *Econometrica*.

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

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*.

Barberis, N. and R. Thaler (2003). A survey of behavioral finance. In *Handbook of the Economics of Finance*.

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*.

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

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

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

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

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

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

Duffie, D., S. Malamud, and G. Manso (2009). Information percolation with equilibrium search dynamics. *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.

French, K. (2008). The cost of active investing. *Journal of Finance*.

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.

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

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

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

Investment Company Institute (2015). *Investment Company Fact Book*. Investment Company Institute. `http://www.icifactbook.org/pdf/2015_factbook.pdf`.

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

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

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

Keynes, J. (1921). *A Treatise on Probability*. Macmillan and Company.

Khandani, A. and A. Lo (2007). What happened to the quants in August 2007? *Journal of Investment Management*.

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.

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

Ozsoylev, H. and J. Walden (2011). Asset pricing in large information networks. *Journal of Economic Theory*.

Pennacchi, G. and M. Rastad (2011). Portfolio allocation for public pension funds. *Journal of Pension Economics and Finance*.

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*.

Stambaugh, R. (2014). Presidential address: Investment noise and trends. *Journal of Finance*.

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

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 Proofs

Definition (Binary String)**.** Let $\{0,1\}^\star = \cup_{n=0,1,2,\ldots}\{0,1\}^n$ denote the set of all 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 $i \in \{0,1\}^\star$

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

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 stock, 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 stock Z is evaluated separately. For example, in the market described by Figure 2, the path described in Equation (11) exists with or without the path described by Equation (12). 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 don't need to check every single subset $\hat{\mathsf{S}} \subseteq \mathsf{S}$ separately. To see which subsets of stocks are connected to stock Z, we can just check which stocks are connected to stock 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). $\square$

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's 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 reasonable instance sizes.

"If cases like this regularly arose in practice, then it would've turned out that we were using the wrong abstraction. But so far, it seems like we're 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 don't 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 an index-fund 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 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's a sharp phase transition in the size of the largest connected component in an Erdös-Rényi random-networks 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 percolation threshold implies that the probability of stock Z being affected by an index-fund rebalancing cascade starting somewhere else in the market is vanishingly small when $\kappa < 1$.

Remark (Connectivity Threshold)**.** There's 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 the giant component and nodes without any edges whatsoever. For our purposes, this connectivity threshold implies that the probability stock Z isn't affected by an index-fund rebalancing cascade starting somewhere else in the market is vanishingly small when $\kappa > \log(S)$.

Proof (Equation 13)**.** Suppose M contains S stocks and was generated using connectivity parameter $\kappa > 0$. If $(s, s') \in \mathsf{S}^2$, then stock s' will be a positive neighbor to stock s with probability ${}^{\kappa}\!/_{S}$. Because the outcome is determined independently for each stock $s' \in \mathsf{S}$, the probability that stock s has exactly n positive neighbors is

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

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 $S \to \infty$

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

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

Proof (Proposition 2.2b)**.** Let $C_s \in \{\texttt{True}, \texttt{False}\}$ be an indicator variable for whether a stock s is connected to the giant component of the random graph induced by M. We can write

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

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 stock $s \in \mathsf{S}$ to stock Z can be evaluated independently. Bayes' rule implies

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

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

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 an arbitrary complexity class, such as NP. We say that Prob is hard with respect to CC if every decision problem in CC can be reduced to Prob.

Definition (Completeness)**.** Let CC denote an arbitrary 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 ($\mathsf{3Sat}$)**.**

- *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)**.** $\mathsf{3Sat}$ *is an* NP*-complete problem.*

Corollary. Let Prob denote any decision problem. If Prob is reducible to $\mathsf{3Sat}$, 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 for each stock $s \in \mathsf{S}$
$$\alpha(k)_{s,t} = 1_{\{(x_{s,t}, \Delta x_{s,t}) = k\}}$$
- For each pair of stocks $(s, s') \in \mathsf{S}^2$ such that $s \neq s'$ define
$$\beta^{+}_{s,s',t} = 1_{\{s' \in \mathsf{Out}^{+}_{s,t}\}}$$
$$\beta^{-}_{s,s',t} = 1_{\{s' \in \mathsf{Out}^{-}_{s,t}\}}$$
- For each pair of stocks $(s, s') \in \mathsf{S}^2$ such that $s \neq s'$ define
$$\gamma^{+}_{s',s,t+1} = 1_{\{s \in \mathsf{In}^{+}_{s',t+1}\}}$$
$$\gamma^{-}_{s',s,t+1} = 1_{\{s \in \mathsf{In}^{-}_{s',t+1}\}}$$
- For each stock $s \in \mathsf{S}$ define
$$\delta^{+}_{s,t+1} = 1_{\{u_{s,t+1} = 1\}}$$
$$\delta^{-}_{s,t+1} = 1_{\{u_{s,t+1} = -1\}}$$

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 can't 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	⊗ $(\bar{\delta}^{+}_{s} \vee \bar{\lambda}^{+}_{s',s} \vee \bar{\lambda}^{-}_{s'',s})$

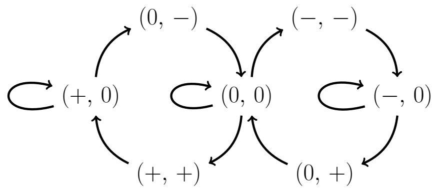


Figure 7: State Diagram. *All possible ways that a single stock could move between the 7 possible values of $(x_{s,t}, \Delta x_{s,t})$ in successive rounds of an index-fund rebalancing cascade. Arrows denote transitions. Loops denote unchanged values in successive rounds.*

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

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

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 Prob $\in$ 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*
$$\mathrm{F} : \{\mathtt{True}, \mathtt{False}\}^N \mapsto \{0,1\}$$
- *Question: Is $\sum_{\mathbf{x} \in \{\mathtt{True},\mathtt{False}\}^N} \mathrm{F}(\mathbf{x}) > 2^{N-1}$?*

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. $\square$

Problem D (2Sat)**.**

- *Instance: A Boolean formula defined over N input variables*
$$\mathrm{F} : \{\mathtt{True}, \mathtt{False}\}^N \mapsto \{0,1\}$$
where no clause contains more than 2 variables.
- *Question: Is there an assignment $\mathbf{x} \in \{\mathtt{True},\mathtt{False}\}^N$ such that $\mathrm{F}(\mathbf{x}) = 1$?*

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

Proof (Proposition 2.4a)**.** If there is no alternation, then stocks only have positive neighbors. So, a stock Z will be affected by an initial shock to the stocks in A if and only if there is a path from stock $s \in \mathsf{A}$ connecting to stock Z. Without alternation, there is no way for two different paths in an index-fund rebalancing cascade to interfere with one another. And, since within a single path, each stock has only O (stock A) or 1 (all other stocks) 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. And, this reduction implies it's solvable in polynomial time (Cook, 1971). □

Proof (Proposition 2.4b)**.** If there are no loops, then there is either a single path from any stock s to stock 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 stock 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 2.4a. Thus, without loops, How is reducible to 2Sat. And, this reduction implies that it's solvable in polynomial time (Cook, 1971). □

Problem E (SmoothHow)**.** *Suppose that the updating rule in Equation* (6) *was changed to the following for some* $\theta \in (0, 1)$*:*

$$\begin{aligned} u_{s,t+1} &= \tfrac{1}{|\mathsf{In}^+_{s,t+1}|+|\mathsf{In}^-_{s,t+1}|} \cdot \left(\textstyle\sum_{s' \in \mathsf{In}^-_{s,t+1}} x_{s',t} - \sum_{s'' \in \mathsf{In}^+_{s,t+1}} x_{s'',t} \right) \\ x_{s,t+1} &= \theta \cdot (x_{s,t} + u_{s,t+1}) \end{aligned}$$

- *Instance: A choice for stock Z; a market structure M; a time $T = \mathrm{Poly}[S]$; a positive constant $\epsilon > 0$; and, the power set $\hat{\mathsf{A}} \subseteq 2^{\mathsf{S}}$ of all ϵ-small sets $\mathsf{A} \subseteq \mathsf{S}$.*
- *Question: Does there exist a $\mathsf{A} \in \hat{\mathsf{A}}$ such that $\mathrm{Effect}_{\mathsf{M},T}(\mathsf{A}, Z) < 0$?*

Proposition 2.4c (Necessity of Thresholds, Restated)**.** *Let i denote an instance of* SmoothHow*. There's a polynomial-time algorithm,* F*, such that for any $\delta > 0$*

$$\textstyle\sum_{|i|=N,\, i \notin \mathsf{Prob}} 1_{\{\mathrm{F}(i)=1\}} < \delta$$

Proof (Proposition 2.4c)**.** Because $\theta < 1$, the effect of a long direct path connecting to stock Z (i.e., where each stock in the path has exactly one incoming neighbor) will decay at an exponential rate. A direct path from stock A to stock Z that involves $(K-1)$ intermediary stocks will have an affect on stock Z proportional to θ^K. And, the effects of any indirect paths (i.e., where each stock in the path has more than one incoming neighbor) will decay even fast due to averaging. Having more than one incoming neighbor presents that possibility that a stock will be hit by both a positive and a negative shock at the same time. So, to get an approximate solution to SmoothHow, just compute the effect of all direct paths connecting to stock Z of length $K = \mathrm{Poly}[S]$. If there exists a path with a negative effect, then answer 'Yes'; otherwise, answer 'No'. □

B Tables

Summary Statistics, ETFs

a) Time Series

	Avg	Sd	1%	25%	50%	75%	99%
#etf	1049	153	863	911	1012	1206	1337
#bmark	876	128	731	768	832	1006	1117
mktCap [$1b]	1291	584	2	984	1420	1683	2283

b) Cross-Section

	Avg	Sd	1%	25%	50%	75%	99%
#stock	241	505	1	29	75	243	2335

Table 1: *Summary statistics for the ETF market. Data is from ETF Global. Sample period runs from January 2010 to December 2015. Panel a) reports monthly aggregates for the entire ETF market.* #`etf`*: number of ETFs in the sample each month.* #`bmark`*: number of different benchmarks reported by these ETFs.* `mktCap`*: total market capitalization of the ETF industry each month in billions of dollars. Panel b) reports cross-sectional statistics for fund-month observations.* #`stock`*: number of stocks held by an ETF in a given month.*

Summary Statistics, Trading

a) Announcements

	Avg	Sd	Min	25%	50%	75%	Max
$\frac{\texttt{\#ancmt}}{\texttt{month}}$	14.73	5.19	4	11	14	19	28
$\frac{\texttt{\#stockZ}}{\texttt{ancmt}}$	1077.02	1151.21	162	456	641	912	4509

b) Stocks

	Avg	Sd	Min	25%	50%	75%	Max
`#etfOwned`	17.39	16.38	1.00	2.87	13.29	27.17	60.08
ln(`etfRebalVlm`)	7.62	2.81	0.65	5.81	8.24	9.68	12.58
`etfOrdImbal`	−0.03	0.14	−0.72	−0.05	−0.02	0.01	0.22
`amihud` [%/$1m]	0.85	3.90	0.00	0.00	0.01	0.07	18.09
`baSpread` [%]	0.01	0.01	0.00	0.00	0.00	0.01	0.05

Table 2: *Summary statistics for trading data. Announcement data is from Thompson Financial. Stock-market data is from CRSP and TAQ. Sample period runs from January 2010 to December 2015. Panel a) reports summary statistics for the M&A announcement data.* $\frac{\texttt{\#ancmt}}{\texttt{month}}$*: number of announcements per month.* $\frac{\texttt{\#stockZ}}{\texttt{ancmt}}$*: number of stocks that are unrelated to the stock named as the target of each M&A announcement. Panel b) reports cross-sectional summary statistics at the stock-month level.* `#etfOwned`*: number of ETFs that own each stock.* ln(`etfRebalVlm`)*: log ETF rebalancing volume.* `etfOrdImbal`*: ETF order imbalance, which ranges from* $[-1, 1]$*.* `amihud`*: Amihud (2002) liquidity measure quoted in units of percent per* \$1m *of trading volume.* `baSpread`*: bid-ask spread quoted as a percent of the closing price.*

Rebalancing Volume, Stock A

	$\ln(\texttt{etfRebalVlm}_{A,t})$ [%]				
	Actual Announcements				Placebo
$1_{\{t=t_A+1\}}$	$22.07^{\star}$ (11.66)	-10.23 (11.78)	$22.14^{\star}$ (11.71)	-10.15 (11.82)	-16.47 (17.41)
$1_{\{t=t_A\}}$	$165.62^{\star\star\star}$ (10.25)	$167.78^{\star\star\star}$ (10.28)	$165.69^{\star\star\star}$ (10.30)	$166.87^{\star\star\star}$ (10.33)	-11.47 (15.42)
$1_{\{t=t_A-1\}}$	$60.66^{\star\star\star}$ (10.52)	$63.94^{\star\star\star}$ (10.35)	$60.93^{\star\star\star}$ (9.82)	$64.19^{\star\star\star}$ (9.56)	-10.53 (19.38)
$1_{\{t=t_A-2\}}$			5.16 (8.33)	6.81 (8.28)	10.98 (17.90)
$1_{\{t=t_A-3\}}$			-2.45 (9.38)	-2.45 (9.35)	-13.28 (16.43)
$1_{\{t=t_A-4\}}$			13.70 (9.71)	14.19 (9.72)	13.21 (21.36)
$1_{\{t=t_A-5\}}$			3.59 (9.15)	5.81 (9.14)	7.45 (15.92)
$\ln(\texttt{vlm}_{A,t-1})$		$12.29^{\star\star\star}$ (0.41)		$12.29^{\star\star\star}$ (0.41)	$12.29^{\star\star\star}$ (0.41)
$\ln(\texttt{vlm}_{A,t-2})$		$3.85^{\star\star\star}$ (0.29)		$3.85^{\star\star\star}$ (0.29)	$3.85^{\star\star\star}$ (0.29)
$\ln(\texttt{vlm}_{A,t-3})$		$3.09^{\star\star\star}$ (0.34)		$3.09^{\star\star\star}$ (0.34)	$3.09^{\star\star\star}$ (0.34)
Month-Year FE	Y	Y	Y	Y	Y
Stock-Specific FE	Y	Y	Y	Y	Y
R^2	57.0%	57.1%	57.0%	57.1%	27.0%
Observations	5,197,276				5,197,276

Table 3: *Effect of initial M&A announcements on ETF rebalancing volume for stock A. For the first 4 columns, the data is panel containing each M&A target in the window $t \in \{t_A - 20, \ldots, t_A + 5\}$. The fifth column uses a new dataset of randomly selected M&A announcement dates for the same target companies. t_A denotes date of M&A announcement for stock A. $\ln(\texttt{vlm}_A)$ is log trading volume for stock A. Table reports results for the regression:* $\ln(\texttt{etfRebal}_{A,t}) = \alpha + \beta \cdot 1_{\{t=t_A-1\}} + \gamma \cdot 1_{\{t=t_A\}} + \delta \cdot 1_{\{t=t_A+1\}} + \cdots + \varepsilon_{A,t}$.

Rebalancing Volume, Stock Z

	$\ln(\texttt{etfRebalVlm}_{Z,t})$ [%]			
$\texttt{afterAncmt}_{A,t}$	$4.98^{\star\star\star}$ (0.28)	$4.85^{\star\star\star}$ (0.28)	$3.27^{\star\star\star}$ (0.39)	$3.02^{\star\star\star}$ (0.39)
$\texttt{manyNbrs}_{A\to Z,t}$			$95.58^{\star\star\star}$ (3.25)	$94.05^{\star\star\star}$ (3.23)
$\texttt{afterAncmt}_{A,t} \times \texttt{manyNbrs}_{A\to Z,t}$			$3.70^{\star\star\star}$ (0.55)	$3.02^{\star\star\star}$ (0.55)
$\ln(\texttt{vlm}_{Z,t})$		$32.38^{\star\star\star}$ (1.72)		$31.45^{\star\star\star}$ (1.70)
Announcement FE	Y	Y	Y	Y
Stock-Specific FE	Y	Y	Y	Y
R^2	52.6%	53.0%	53.3%	53.6%
Observations	14,736,786		14,736,786	

Table 4: *Effect of initial M&A announcements on ETF rebalancing volume for stock Z. Data is panel containing each stock Z that is unrelated to the target stock A in the window $t \in \{t_A - 20, \ldots, t_A + 5\}$. For stock Z to be unrelated to a particular stock A, it has to be twice removed in the network of ETF holdings. It can't have been recently held by any ETF that also recently held stock A. And, it can't have been held by any ETF that also held a stock that was held by another ETF that held stock A. i.e., the chain has to be $A - B - C - Z$ or longer. $\texttt{afterAncmt}_A$ is an indicator variable for the 5 days following the announcement of stock A as an M&A target. $\texttt{manyNbrs}_{A\to Z}$ is an indicator variable for stock Z having an above-median number of neighbors relative to stock A's M&A announcement. $\ln(\texttt{vlm}_Z)$ is log trading volume for stock Z. Table reports results for the regression:* $\ln(\texttt{etfRebalVlm}_{Z,t}) = \alpha + \beta\cdot\texttt{afterAncmt}_{A,t} + \gamma\cdot\texttt{manyNbrs}_{A\to Z,t} + \delta\cdot\{\texttt{afterAncmt}_{A,t} \times \texttt{manyNbrs}_{A\to Z,t}\} + \cdots + \varepsilon_{A\to Z,t}$.

Order Imbalance, Stock Z

	$\texttt{etfOrdImbal}_{Z,t}$ [bps]			
$\texttt{afterAncmt}_{A,t}$	0.75^{***} (0.11)	0.74^{***} (0.11)	0.63^{***} (0.16)	0.63^{***} (0.16)
$\texttt{manyNbrs}_{A\to Z,t}$			-0.92^{***} (0.10)	-0.87^{***} (0.10)
$\texttt{afterAncmt}_{A,t} \times \texttt{manyNbrs}_{A\to Z,t}$			0.24 (0.21)	0.22 (0.21)
$\ln(\texttt{vlm}_{Z,t})$		-1.25^{***} (0.08)		-1.24^{***} (0.08)
Announcement FE	Y	Y	Y	Y
Stock-Specific FE	Y	Y	Y	Y
R^2	1.4%	1.4%	1.4%	1.4%
Observations	13,755,851		13,755,851	

Table 5: *Effect of initial M&A announcements on ETF order imbalance for stock Z. Data is panel containing each stock Z that is unrelated to the target stock A in the window $t \in \{t_A - 20, \ldots, t_A + 5\}$. For stock Z to be unrelated to a particular stock A, it has to be twice removed in the network of ETF holdings. It can't have been recently held by any ETF that also recently held stock A. And, it can't have been held by any ETF that also held a stock that was held by another ETF that held stock A. i.e., the chain has to be $A - B - C - Z$ or longer.* $\texttt{afterAncmt}_A$ *is an indicator variable for the 5 days following the announcement of stock A as an M&A target.* $\texttt{manyNbrs}_{A\to Z}$ *is an indicator variable for stock Z having an above-median number of neighbors relative to stock A's M&A announcement.* $\ln(\texttt{vlm}_Z)$ *is log trading volume for stock Z. Table reports results for the regression:* $\texttt{etfOrdImbal}_{Z,t} = \alpha + \beta \cdot \texttt{afterAncmt}_{A,t} + \gamma \cdot \texttt{manyNbrs}_{A\to Z,t} + \delta \cdot \{\texttt{afterAncmt}_{A,t} \times \texttt{manyNbrs}_{A\to Z,t}\} + \cdots + \varepsilon_{A\to Z,t}$.

Liquidity Measures, Stock Z

	$\texttt{amihud}_{Z,t}$ [%/\$1m]		$\texttt{baSpread}_{Z,t}$ [bps]	
$\texttt{afterAncmt}_{A,t}$	0.80 (0.63)	0.84 (1.09)	$0.17^{\star\star\star}$ (0.06)	$0.18^{\star}$ (0.10)
$\texttt{manyNbrs}_{A\to Z,t}$		$-4.61^{\star\star\star}$ (1.51)		$-5.31^{\star\star\star}$ (0.40)
$\texttt{afterAncmt}_{A,t} \times \texttt{manyNbrs}_{A\to Z,t}$		-0.09 (1.41)		-0.04 (0.12)
Announcement FE	Y	Y	Y	Y
Stock-Specific FE	Y	Y	Y	Y
R^2	6.7%	6.7%	52.8%	52.9%
Observations	14,736,786		14,736,786	

Table 6: *Effect of initial M&A announcements on liquidity for stock Z. Data is panel containing each stock Z that is unrelated to the target stock A in the window $t \in \{t_A - 20, \ldots, t_A+5\}$. For stock Z to be unrelated to a particular stock A, it has to be twice removed in the network of ETF holdings. It can't have been recently held by any ETF that also recently held stock A. And, it can't have been held by any ETF that also held a stock that was held by another ETF that held stock A. i.e., the chain has to be $A-B-C-Z$ or longer.* $\texttt{amihud}_Z$ *is Amihud (2002) illiquidity measure in units of % per million dollars.* $\texttt{baSpread}_Z$ *is bid-ask spread as a fraction of closing price.* $\texttt{afterAncmt}_A$ *is an indicator variable for the 5 days following the announcement of stock A as an M&A target.* $\texttt{manyNbrs}_{A\to Z}$ *is an indicator variable for stock Z having an above-median number of neighbors relative to stock A's M&A announcement. For $\texttt{y} \in \{\texttt{amihud}, \texttt{baSpread}\}$, table reports results for the regression:* $\texttt{y}_{Z,t} = \alpha+\beta\cdot\texttt{afterAncmt}_{A,t}+\gamma\cdot\texttt{manyNbrs}_{A\to Z,t}+\delta\cdot\{\texttt{afterAncmt}_{A,t}\times\texttt{manyNbrs}_{A\to Z,t}\}+\cdots+\varepsilon_{A\to Z,t}$.

Unconditional Panel Regression

	$\ln(\texttt{etfRebalVlm}_{s,t})$ [%]		$\texttt{amihud}_{s,t}$ [%/\$100m]		$\texttt{baSpread}_{s,t}$ [bps×100]	
$\texttt{\#nhbrs}_{s,t}$	$0.12^{\star\star\star}$ (0.01)	$0.06^{\star\star\star}$ (0.01)	$-1.47^{\star\star\star}$ (0.31)	$-0.63^{\star\star}$ (0.27)	$-0.88^{\star\star\star}$ (0.08)	$-0.25^{\star\star\star}$ (0.06)
Month-Year FE	Y		Y		Y	
#etf-Specific FE		Y		Y		Y
Stock-Specific FE	Y	Y	Y	Y	Y	Y
R^2	60.6%	43.4%	6.5%	6.5%	53.6%	53.7%
Observations	4,915,505		4,966,292		4,986,730	

Table 7: *Unconditional relationship between number of neighbors and a stock's ETF rebalancing volume and liquidity. Data consists of a panel containing all stock-month observations in our sample.* $\#\texttt{nbrs}_s$ *is the number of neighboring stocks to stock* s *in the ETF rebalancing-rule network.* $\texttt{amihud}_s$ *is Amihud (2002) illiquidity measure in units of* % *per* 100 *million dollars.* $\texttt{baSpread}_s$ *is bid-ask spread as a fraction of closing price times* 100*.* #etf*-specific fixed effects denote indicator variables for the number of ETFs that hold a given stock in a given month. For* $\texttt{y} \in \{\ln(\texttt{etfRebalVlm}), \texttt{amihud}, \texttt{baSpread}\}$*, table reports results for the regressions* $\texttt{y}_{s,t} = \alpha + \beta \cdot \#\texttt{nbrs}_{s,t} + \cdots + \varepsilon_{s,t}$.