The Sound Of Many Funds Rebalancing[*]

Alex Chinco[†] and Vyacheslav Fos[‡]

December 24, 2018

Abstract

This paper proposes that complexity generates noise in financial markets. There is a large number of funds following a wide variety of different trading rules. And, because it's too computationally complex to predict how these trading rules will interact with one another, a stock's demand can appear random even if you yourself are fully rational. We first model a particular kind of trading rule—index-fund rebalancing—to show how complexity can generate demand noise. In the model, it's easy to predict *if* a stock will be involved in an index-fund rebalancing cascade, but it's computationally infeasible to predict *how* the stock's demand will be affected (buy? or sell?). As a result, traders treat the demand coming from index-fund rebalancing cascades as noise. We then analyze the rebalancing activity of a particular kind of index fund—exchange-traded funds (ETFs)—to give empirical evidence that complexity actually does generate demand noise in real-world financial markets. We document that ETF rebalancing cascades transmit economically large demand shocks that are statistically unpredictable.

JEL CLASSIFICATION. G00, G02, G14

KEYWORDS. Noise, Indexing, Thresholds, Complexity

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 about the stock. 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, there are always people buying and selling stock Z for erratic non-fundamental reasons. So, when you try to buy a share, other traders won't jump directly to the conclusion that you've uncovered good news. Your buy order might just be some more random noise. And, it's this convenient cover story that allows you to both trade on and profit from your discovery. It's this plausible alibi that "makes financial markets possible (Black, 1985)".

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

The standard answers to these questions are that i) demand noise comes from individual investors and ii) individual investors' demand looks erratic and unrelated to fundamentals because they are just plain bad traders. We know that individual investors suffer from all sorts of behavioral biases when they trade (Barberis and Thaler, 2003), and we also know that they trade far too often (Barber and Odean, 2000). So, these are the standard answers for a reason. It's clear that individual investors can generate demand noise.

But, are they the only source? It seems unlikely. After all, 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, by June 2017, only "10% of trading was done by traditional, 'discretionary' traders, as opposed to systematic rules-based ones."[1] Yet, in spite of this drop in the importance of individual investors, there's been no corresponding decline in trading volume.

So, we propose an alternative noise-generating mechanism: complexity. Financial markets contain a large number of funds following a wide variety of different trading rules. And, if it's too computationally complex to predict how these various trading rules will interact with one another, then the resulting demand will appear random even if you yourself are fully rational. We start by modeling a particular kind of trading rule—index-fund rebalancing—to show theoretically how complexity can generate noise. We then study the rebalancing activity of a particular kind of index fund—exchange-traded funds (ETFs)—to give empirical evidence that complexity actually does generate noise in real-world financial markets.

---

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

Theoretical Model. As retail investing has 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)." These new *index-like* funds have not been created in Jack Bogle's image, however. Many choose their holdings based on custom threshold-based rules. For instance, the PowerShares S&P 500 Low-Volatility ETF [SPLV] tracks a benchmark consisting of the 100 lowest volatility S&P 500 stocks. This benchmark involves a threshold 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 of these 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 using a different threshold 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* an unrelated stock Z will be affected by one of these index-fund rebalancing cascades, the problem of determining *how* this stock Z will be affected (buy? or sell?) is computationally intractable. In fact, it's NP hard. As a result, even if every underlying index fund is following a completely deterministic trading rule, fully rational traders will still treat the direction of the demand shock coming from this index-fund rebalancing cascade as a random coin flip—i.e., as demand noise.

This model delivers two key insights. The first relates to why index-fund rebalancing cascades are complex. In the past, when financial economists saw a complex market outcome, they went looking for some complex market input, such as an abstruse financial footnote (Loughran and McDonald, 2011) or a convoluted derivative payout (Arora et al., 2011). But, the model shows that market complexity is much more pervasive. It can emerge even when all these complex inputs have been stripped away; it can emerge even in a market involving only index funds making simple deterministic rebalancing decisions. The second insight concerns the nature of the resulting complexity. The model doesn't just hint that some traders might be confused by the complexity of index-fund rebalancing cascades. You don't need a model to see that. Instead, the model shows that no trader will be able to predict whether an index-fund rebalancing cascade will result in buy or sell orders for an unrelated stock Z. Thus, even the simplest decisions, such as those involved in index-fund rebalancing, can generate demand shocks that look provably random to all traders, which suggests that computational complexity is an important source of demand noise in modern financial markets.

Empirical Evidence. Having outlined the model, we next provide evidence that computational complexity actually does generate demand noise in the real world. We do this by studying the end-of-day holdings of a particular kind of index fund—exchange-traded funds (ETFs)—from January 2011 to December 2017 using data from ETF Global. We net-out changes in ETF holdings due to creations and redemptions since these trades are executed as in-kind transfers for tax reasons (Madhavan, 2016; Ben-David et al., 2017). We also restrict our data to only include ETFs that rebalance more than once a quarter. So, when you look at our results, you should have in mind the PowerShares S&P 500 Low-Volatility ETF rather than the SPDR S&P 500 ETF. It's true that ETFs which rebalance more often also tend to be smaller than ETFs which track broad value-weighted market indexes that rebalance infrequently. But, the rebalancing activity of these smaller ETFs still matters because they tend to do all their trading during the final 20-to-30 minutes of the trading day.[2]

In the model, we study index-fund rebalancing cascades that stem from an initial shock to some stock A. So, in our empirical analysis, we have to make a decision about which initial shocks to use. To this end, we study ETF rebalancing activity in the days around an M&A announcement, referring to the target of the M&A announcement as stock A. 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—Wednesday vs. Thursday—may as well be. Our data on M&A announcements comes from Thomson Financial.

We then look for evidence that ETF rebalancing cascades transmit the effects of these initial M&A-announcement shocks to other stock Zs that are unrelated to the M&A target, stock A. For stock A and stock Z to be unrelated, they must be twice removed in the network of ETF holdings at the time of the M&A announcement. Stock Z can't have been recently held by any ETF that also recently held stock A. And, if stock A and stock B were both recently held by the same ETF, then stock Z also can't have been recently held by any ETF that recently held stock B. In other words, the chain of ETF rebalancing decisions connecting stock A and stock Z must be $A \to B \to C \to Z$ or longer. Because there are smart-beta ETFs tracking things like large-cap, value, and industry, this double-separation criteria also implies that stock A and stock Z have dissimilar factor exposures and firm characteristics.

The model predicts that i) an unrelated stock Z that's on the cusp of many ETF rebalancing thresholds is more likely to be hit by an ETF rebalancing cascade than an unrelated stock Z that's on the cusp of few rebalancing thresholds; however, ii) it shouldn't be possible to predict the direction of any resulting demand shock. To test these predictions, we split

---

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

each set of stock Zs that's unrelated to the target of an M&A announcement, stock A, into two subsets: those that are on the cusp of rebalancing for an above-median number of ETFs, and those that aren't. After verifying that these two groups of stocks do not display different pre-trends in the days leading up to each M&A announcement, we show that ETF rebalancing volume is 2.06% higher for the above-median stock Zs than for the below-median stock Zs in the five days immediately after an M&A announcement. But, we also show that this increase in ETF rebalancing volume 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. We run two placebo tests to further bolster these results. The first involves using randomly assigned announcement dates while the second involves focusing on a subset of ETFs that rarely rebalances. In addition, we show that the unpredictable demand coming from ETF rebalancing cascades affects the prices of the underlying stocks. Large positive demand shocks to price increases; whereas, lots of selling pressure leads to price decreases.

What's more, there are several reasons why this point estimate represents a lower bound on the amount of demand noise generated by computational complexity. First, not all ETF rebalancing cascades begin with an initial M&A announcement. We estimate that a 1σ increase in a stock's exposure to ETF rebalancing cascades more generally—i.e., not just those starting with an M&A announcement—is associated with a 0.4σ increase in ETF rebalancing volume. Second, not all rebalancing cascades involve ETFs. Our theoretical analysis applies to any situation where there's a large group of funds following a wide variety of threshold-based rebalancing rules. For other examples, 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 pension funds with mandates of the form *'15% of our assets must be held in asset class X'* (Pennacchi and Rastad, 2011). The interactions of these funds' trading rules can also generate demand noise. Third, rebalancing cascades are not the only source of complexity in financial markets. Financial markets are complex for all sorts of reasons. We focus our attention on rebalancing cascades because we can theoretically prove that it's computationally infeasible to predict how they will affect a stock's demand and then provide empirical evidence that this complexity results in demand noise.

1.1 Related Literature

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

Noise. Noise plays a central role 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). The
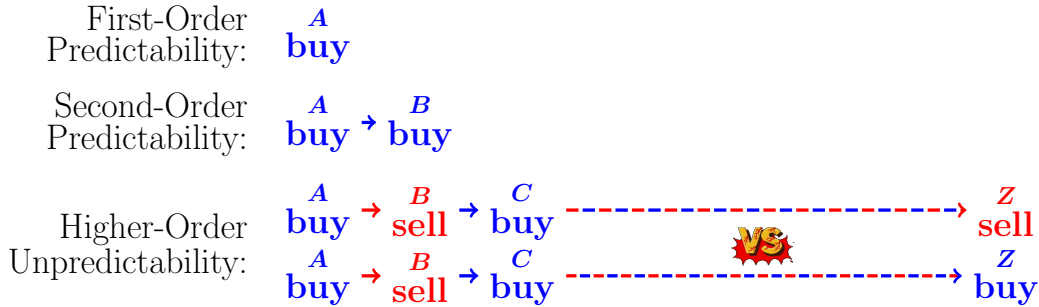
First-Order Predictability:	*A* **buy**				
Second-Order Predictability:	*A* **buy** →	*B* **buy**			
Higher-Order Unpredictability:	*A* **buy** →	*B* **sell** →	*C* **buy** -----------------→	VS	*Z* **sell**
	A **buy** →	*B* **sell** →	*C* **buy** -----------------→		*Z* **buy**

Figure 1. Existing Papers on Indexing. Existing papers on the effects of index-linked investing can be classified into one of two subgroups. The first studies how trading due to index inclusion can directly affect a stock (Row 1). The second studies how stock *A*'s inclusion in an index will affect its relationship with stock *B* in a predictable fashion (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 an unrelated stock *Z* that's several steps removed (Row 3). A seemingly innocuous change in the buy-sell-buy-sell sequence connecting stock *A* to stock *Z*—i.e., the slight difference in the two dashed lines—can completely change how a rebalancing cascade will affect stock *Z*—i.e., whether the cascade result in buy or sell orders for stock *Z*.

key contribution of this paper is to propose an explanation for where this demand noise comes from that does not rely on individual investors behaving randomly. What's more, the mechanism proposed in this paper differs from other approaches to doing this, such as hedging demand, because it makes quantitative predictions about which assets should have more/less demand noise.

Indexing. Our paper also relates to work on index-linked investing (Wurgler, 2010). This literature can be classified into two subgroups. The first group of papers studies how index inclusion directly affects the underlying stocks in a predictable way (Figure 1, Row 1). For instance, papers such as Bessembinder (2015), Shum et al. (2015), Bai et al. (2015), and Ivanov and Lenkey (2018) all study the predictable effects of ETF rebalancing decisions on stock *A*. And, Chang et al. (2014) shows how getting added to the Russell 2000 predictably results in further price increases for stock *A*. For other examples, see Ben-David et al. (2017), Bessembinder et al. (2016), Brown et al. (2016), and Israeli et al. (2017). The second group studies how stock *A*'s inclusion in an index will affect its relationship with some other stock *B* in a predictable fashion (Figure 1, Row 2). For instance, Barberis et al. (2005) shows that a stock's beta with the S&P 500 jumps sharply after index inclusion. For more examples, see Greenwood and Thesmar (2011), Vayanos and Woolley (2013), and Anton and Polk (2014). By contrast, this paper focuses on the unpredictable consequences of stock *A*'s index inclusion, not for stock *A* or stock *B*, but for a completely unrelated stock *Z* (Figure 1, Row 3).

Thresholds. In addition, this paper connects to a broad behavioral-economics literature studying thresholds. People use heuristic threshold-based decision rules in all sorts of different contexts (Gabaix, 2014). The existing literature typically measures the cost of using a heuristic rule in terms of its *expected*-utility loss (Bernheim and Rangel, 2009). Whereas, we look at how simple decision rules can affect demand *volatility*. In other words, we show that the interaction of many different heuristics can have important aggregate effects that a researcher would miss by looking only at the consequences of a single agent's decision rule.

Complexity. Finally, our paper adds to a line of literature studying complexity and chaos in financial markets (e.g., Baumol and Benhabib, 1989; Frank and Stengos, 1989; Scheinkman and LeBaron, 1989; Hsieh, 1991; Rosser, 1999). The current paper is clearly inspired by this ambitious earlier work. But, we would also like to point out an important distinction. These earlier papers use complexity to show that economic phenomena, such as financial crises, are fundamentally unpredictable. Notice that most of the papers cited above were published en masse following the 1987 crash. By contrast, the current paper is using complexity constructively. We show how computational complexity can generate the demand noise which makes financial markets possible and show where this noise will be the loudest, giving a new way to empirically test existing asset-pricing models.

2 Theoretical Model

This paper proposes that complexity generates noise in financial markets. 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 individual index fund is following a simple, deterministic, rebalancing rule. The key insights from the model are that market outcomes can be complex even if individual agents are behaving in a simple way and that this complexity is provably hard to untangle.

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 via their rebalancing rules. These rebalancing rules are going to be extremely simple. And, that's the precisely point. One of the goals of the model is to show how complexity can generate noise

even if individual agents are following extremely simple decision rules. Using a more realistic set of rules will just make predicting the effect of a rebalancing cascade even harder.

Network. Imagine a market where index-fund rebalancing rules define a network over a set of stocks $\mathsf{S} = \{1, 2, \ldots, S\}$. 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:

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

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:

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

A market's structure is the set of all neighbors for each stock, $\mathsf{M} = \{(\mathsf{N}_1^+ \| \mathsf{N}_1^-), \ldots, (\mathsf{N}_S^+ \| \mathsf{N}_S^-)\}$.

Distortion. We want to analyze how this index-fund rebalancing network propagates shocks through the market in discrete rounds indexed by $t = 0, 1, 2, \ldots$ And, for this to happen, it's important that demand curves slope down (Shleifer, 1986). 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. This assumption is consistent both with trader descriptions and current academic research (Ben-David et al., 2017). More and more people are talking about how ETF rebalancing "influences trading in individual stocks."[3] And, there's a lot of overlap between index-fund portfolios. A stock is often involved in numerous ETF benchmarks: "active beta, momentum, dividend growth, deep value, quality, and total earnings."[4]

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:

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

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 a stock's *characteristics* rather than its price. For example, if a large-cap fund decides to buy a stock, then this additional buying pressure might increase

---

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

the stock's volatility enough to force a second low-vol fund to exit its position in the stock.

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 completely deterministic three-step process. First, STEP 1 involves identifying the all stocks that will be affected at time $(t + 1)$ by index-fund rebalancing decisions made at time t:

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

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

$\mathsf{Out}^-_{s,t}$ 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 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.

Next, 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 due to decisions made at time t:

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

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 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 the net effect of this collection of incoming links will distort the characteristics of each stock at time $(t + 1)$:

$$u_{s,t+1} = 1_{\{|\mathsf{In}^+_{s,t+1}| > |\mathsf{In}^-_{s,t+1}|\}} - 1_{\{|\mathsf{In}^+_{s,t+1}| < |\mathsf{In}^-_{s,t+1}|\}}$$
$$x_{s,t+1} = \text{Sign}[x_{s,t} + u_{s,t+1}] \tag{1}$$

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, then stock s will realize a negative shock.

9

Cascades. A rebalancing cascade starts at time $t = 0$ with all stocks at their default levels:

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

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

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

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 $S \to \infty$.

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

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

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 $\text{Cascade}_{\mathsf{M},T}(\mathsf{A})$. The positive-initial-shock convention is without loss of generality. Notice that how description of an index-fund rebalancing cascade suggests a second interpretation for the symbol M. The symbol M doesn't just represent a description of the index-fund rebalancing rules in a particular market. It also represents a description of a machine that computes the effects of any resulting 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.
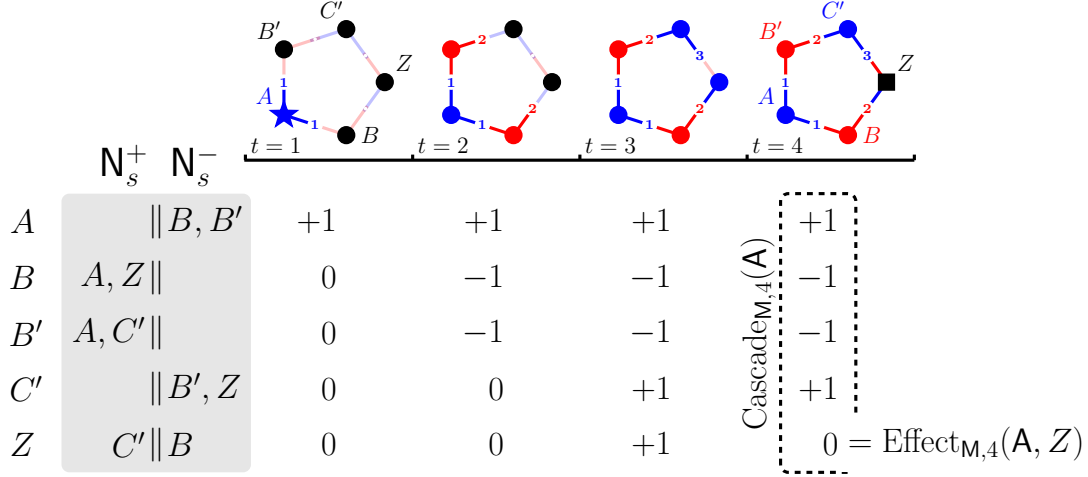
Figure 2. An Example. An index-fund rebalancing cascade involving 5 stocks that starts with a positive shock to stock A, $\mathsf{A} = \{A\}$. Grey box depicts market structure, M. Columns to the right depict state of each stock, $x_{s,t}$, at times $t = 1, 2, 3, 4$. Diagrams above depict propagation of shock through the market. Dots denote stocks. Red dot is $x_{s,t} = +1$; blue dot is $x_{s,t} = -1$; black dot is $x_{s,t} = 0$. Dashed box reports result of index-fund rebalancing cascade at time $t = 4$, $\mathrm{Cascade}_{\mathsf{M},4}(\mathsf{A})$. Notice that cascade has positive effect on stock Z in round $t = 3$, $\mathrm{Effect}_{\mathsf{M},3}(\mathsf{A}, Z) = +1$. But, in round $t = 4$, its net effect on stock Z reverts to $\mathrm{Effect}_{\mathsf{M},4}(\mathsf{A}, Z) = 0$.

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? It turns out that the answer to this question is: *'Not very hard.'* And, we now explore why this is the case.

Decision Problem. Figuring out whether an index-fund rebalancing cascade starting with stock A might affect stock Z means finding at least one path connecting stock A to stock Z in the ETF rebalancing network. We define a K-path connecting stock A to stock Z as 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 for each stock $k \in \{2, \ldots, K-1\}$ we have that
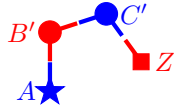
$$
s_k \in \begin{cases} \mathsf{N}^+_{s_{k-1}} & k \text{ odd} \\ \mathsf{N}^-_{s_{k-1}} & k \text{ even} \end{cases} \qquad \text{for all } k \in \{2, \ldots, K\}
$$

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:

$$
\big\{(\emptyset \| \{B\}), (\{A,Z\} \| \emptyset), (\emptyset \| \{B\})\big\} \tag{2}
$$
$$
\underset{\text{Stock } A \;\rightarrow}{} \quad \underset{\text{Stock } B}{} \quad \underset{\rightarrow \text{Stock } Z}{}
$$

11

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

$$\big\{(\emptyset\|\{B'\}),(\{A,C'\}\|\emptyset),(\emptyset\|\{B',Z\}),(\{C'\}\|\emptyset)\big\} \tag{3}$$
$$\underset{\textbf{Stock } A \;\rightarrow\; \textbf{Stock } B' \;\rightarrow\; \textbf{Stock } C' \;\rightarrow\; \textbf{Stock } Z}{}$$

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 $K \leq T$?*

The set If denotes the set of instances where the answer is *'Yes'*. Solving the 'If?' problem means deciding whether $(Z,\mathsf{M},T,\hat{\mathsf{S}}) \in$ If. So, when it's the case that $(Z,\mathsf{M},T,\hat{\mathsf{S}}) \in$ If, 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. In other words, it's easy to determine which stock As have the potential to trigger an index-fund rebalancing cascade that affecting 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 $(Z,\mathsf{M},T,\hat{\mathsf{S}}) \in$ 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)

$$N_s^+ \sim \mathrm{Poisson}(\kappa, S) \tag{4}$$

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 fact that there are "now more indexes than stocks"[5] suggests that the typical stock is on the cusp of more than one rebalancing threshold, $\kappa > 1$.

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*

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

denotes the number of stocks with a K-path to stock Z for some $K \leq T$, then $\mathrm{E}[\hat{\mathsf{S}}_{\mathrm{max}}(Z, \mathsf{M}, T)]$ 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. All you have to do is count stock Z's immediate neighbors. We will make use of this fact in our empirical analysis below.

We want to emphasize, however, that our main results are not about the emergence of a giant component in a random graph when the connectivity parameter crosses $\kappa = 1$. In fact, Proposition 2.2b takes the existence of the giant connected component as given; it assumes $\kappa > 1$. And, this is a reasonable assumption to make in our setting since, as discussed above, modern financial markets contain more indexes than stocks. Our main point is that the interactions between a large number of heterogeneous rebalancing rules in this giant component can create seemingly random demand shocks.

2.3 'How?' Problem

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

Some Intuition. What does it mean to say that *if* is an easier question than *how*? To build some intuition, let's start by looking at Figure 3. Each row depicts a single market with $S = 25$ stocks. Here's the exercise we have in mind. First, examine the left panel, which depicts the index-fund rebalancing rules that define each market, in each row. Then, ask yourself: i) *'Will stock Z, which is denoted by the large black square with a question mark in*

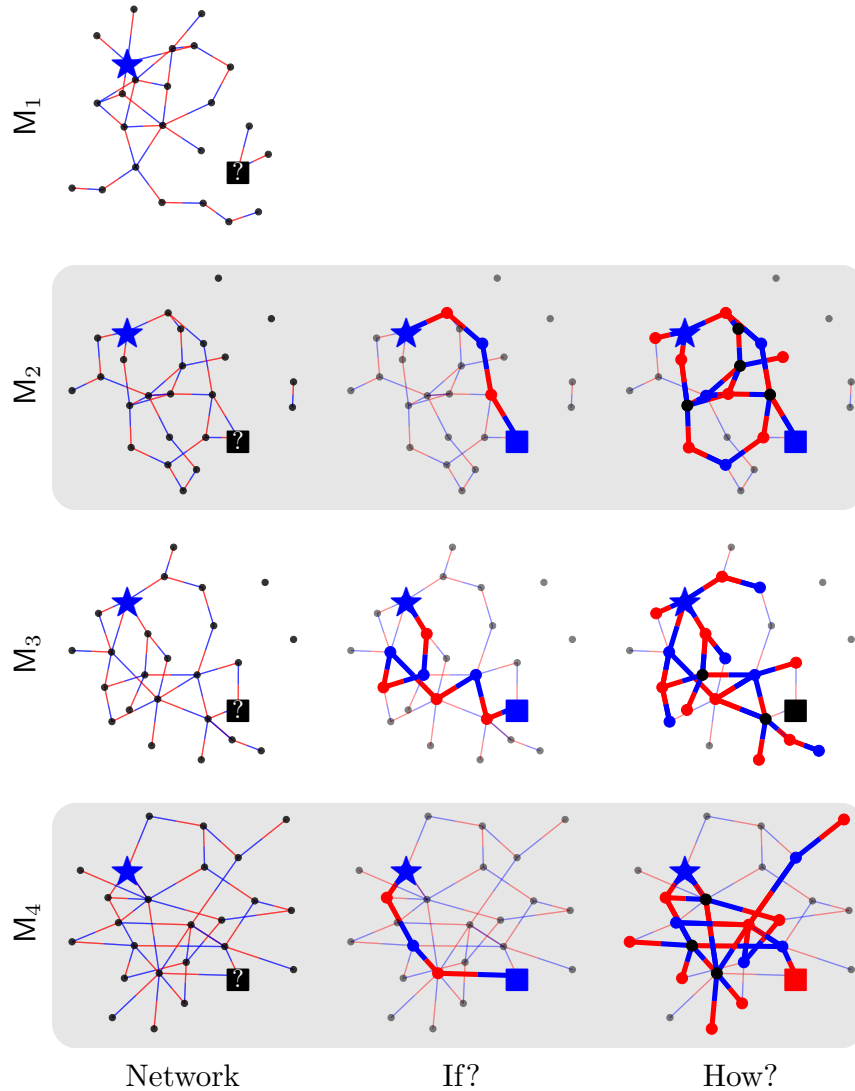[5]Bloomberg. 5/12/2017. *There Are Now More Indexes Than Stocks.*

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.

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

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 markets M_2, M_3, and M_4 (rows two through four) but not in market M_1 (row one). 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 an easy polynomial-time solution. All you have to do is find a break in the chain.

On the other hand, you can also immediately *see* how hard it is to answer 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. The middle column shows that markets M_2, M_3, and M_4 (rows two through four) all have individual paths connecting stock A to stock Z that end with a positive demand shock for stock Z. But, the right column shows that the net effect of the entire index-fund rebalancing cascade in each of these markets only agrees with this naïve prediction for market M_2 in the second row. Even though there are individual paths from stock A to stock Z that would result in a positive demand shock for stock Z in markets M_3 and M_4, neither rebalancing cascade results in a positive demand shock for stock Z on net.

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

The set How denotes the set of instances where the answer is *'Yes'*. Here's what this means 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? Solving How means answering this question.

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 an instance of How is 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'll only take you a second to verify your guess about the answer for 31-down once you see 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 it belongs to NP and every other decision problem in NP can be reduced to it.

Root of the Problem. Figure 4 illustrates precisely 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, as in M_0, then the secondary path doesn't matter. If there is a 1-path connecting stock A to stock D, as in M_1, then 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, as in M_2, then the secondary path won't matter once again. And, if there's a 3-path connecting stock A to stock D, as in M_3, then stock Z will be positively affected by the index-fund rebalancing cascade even though stock D will be unaffected. Thus, tiny changes in the structure of an index-fund rebalancing network can cause an index-fund rebalancing cascade to affect stock Z in completely different ways.

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. And, 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

---

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

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. Never let the perfect be the enemy of the good. So, what if you only tried 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 $T = \text{Poly}[S]$; 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* $\sum_{\mathsf{A} \in \hat{\mathsf{A}}} 1_{\{\text{Effect}_{\mathsf{M},T}(\mathsf{A}, Z) = +1\}} > |\hat{\mathsf{A}}|/2$?

Compared to How, the MajorityHow problem 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 how some stock Z will be affected by the majority of index-fund rebalancing cascade that might possibly occur. For a particular market structure, will more than half of all possible index-fund rebalancing cascades result in buy orders stock Z?

And, at first, the MajorityHow problem seems much easier 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. Moreover, a polynomial-time solution to MajorityHow would imply P = NP.

2.4 Key Ingredients

We've just seen that predicting how an index-fund rebalancing cascade will affect a stock's demand with accuracy better than a coin flip is an NP-hard problem, even if you are fully rational and the index funds involved are following simple deterministic rebalancing rules. As a result, traders have no choice but to treat the demand shock coming from index-fund
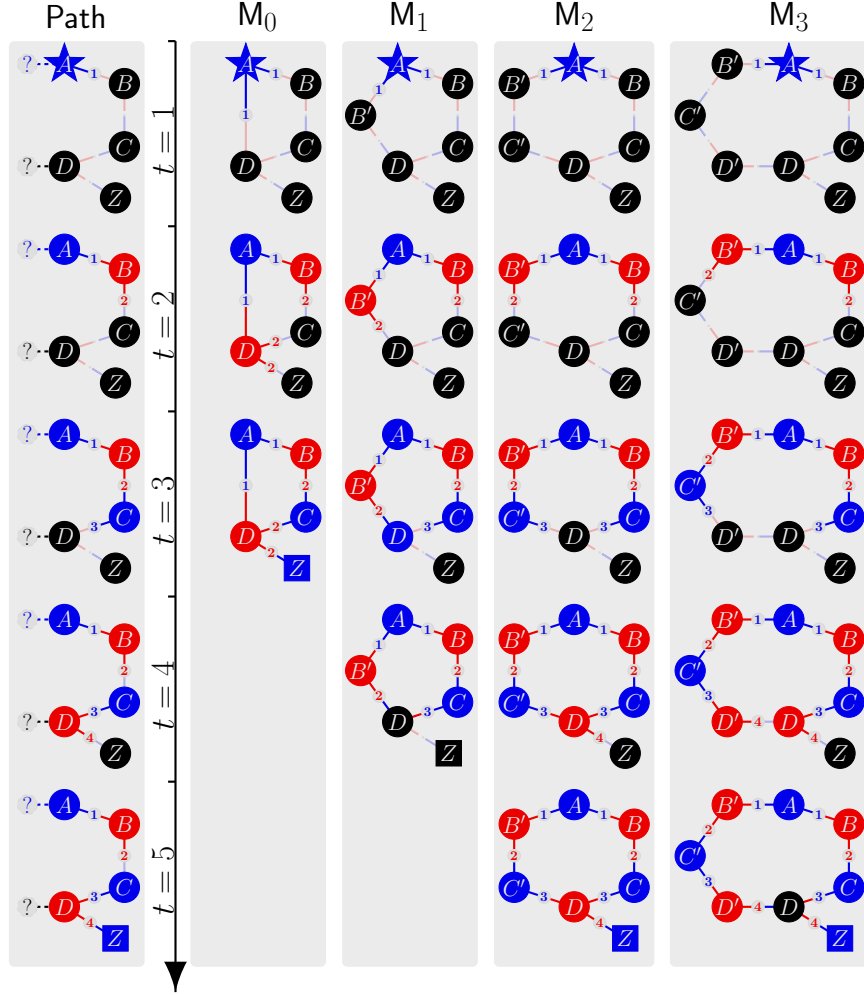
Figure 4. Root of the Problem. Each vertical gray region denotes a separate sequence events, which starts at the top and ends at the bottom. Each node denotes a stock. Node color denotes effect of cascade: blue=positive, red=negative, black=no effect. Star: initial shock to stock A. Square: final effect 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 left column depicts a path connecting stock A to stock Z via stocks B, C, and D that results in a buy order for stock Z when ignoring the other connections that stocks A and D have. The right four columns then illustrate how tiny changes in how stocks A and D are connected to the rest of the network can alter the effect of a cascade starting with stock A on stock Z. The cascade will result in buy orders for stock Z in markets M_0, M_2, and M_3 but not in market M_1. Moreover, you can't tell which situation you are in just by looking at stock Z's immediate neighbors. While stock D realizes no net effect in both market M_1 and market M_3, the cascade only results in a positive net demand shock for stock Z in market M_3.

rebalancing cascades as noise. We now describe three key features of index-fund rebalancing cascades that make them so hard to predict.

Financial economists have long intuited a connection between apparent randomness and complexity. For example, 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?'* may as well be a coin toss. However, in the past, this intuitive connection was just that—an intuition. When a financial economist encountered a complex process with seemingly random output, he had no way of knowing if these data would also look like random noise to a trader with more computational resources. Our theoretical model makes this connection concrete by proving that the results of an index-fund rebalancing cascade appear random for all traders. And, by showing precisely why no one can predict the effects of an index-fund rebalancing cascade, we make it possible for researchers to identify other situations where the same logic holds.

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—the decision is about whether to sell or do nothing at all. There is no alternation involved. And, 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; Brunnermeier and Pedersen, 2008).

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 highlighted in Figures 2 and 4. To illustrate, think about what would happen if every stock in the market had no more than two neighbors. In this setting, there's no way for a single stock to be affected by a rebalancing cascade more than once. So, 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 the number's odd, the shock to stock Z will

be positive, like in Equation (2); if it's even, the shock will be negative, like in Equation (3).

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 holding a single stock. There's nothing stopping 20 different smart-beta ETFs from holding the same company at the same time for different reasons.[7] Thus, the associated collection 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 exponentially weaker at longer and longer horizons. So, without thresholds, you can approximate a cascade's net effect on stock Z with the effect of the shortest path.

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.

We know that index funds use threshold-based rebalancing rules. In fact, this is how most ETFs 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, our goal is to point out one natural consequence of these sorts of trading rules: demand noise.

No-Trade Theorem. Finally, you might be wondering why the classic Milgrom and Stokey (1982) no-trade theorem doesn't 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

---

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

start out with common priors and then one of them gets a private signal. Next, they prove that, if this lone trader acts on his private signal using a simple deterministic trading rule, 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 breaks 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.

3 Empirical Evidence

We've just seen a model showing how it's theoretically possible for complexity to generate demand noise. But, is there any evidence that complexity actually generates demand noise in real-world financial markets? To answer this question, we analyze data on a particular kind of index fund—exchange-traded funds (ETFs)—following a particular kind of initial shock— M&A announcements. Throughout our analysis, we will use `teletype font` to indicate variables that are either used in a regression or get reported as summary statistics.

3.1 Exchange-Traded Funds

The model characterizes index-fund rebalancing cascades that occur following an initial shock to some stock A. So, to bring the model to the data, we need to choose a group of index funds to study as well as a collection of initial shocks. We now address the first of these choices. We use data on the end-of-day positions of U.S. exchange-traded funds (ETFs) from January 1st, 2011 to December 31st, 2017. Our data come from ETF Global. In our analysis, we only consider ETFs that rebalance more than once a quarter. Here are the three reasons why we focus our empirical attention on this particular market.

Reason #1: Apt Description. First, our theoretical model describes a market populated by a large group of index funds following a heterogeneous collection of overlapping benchmark indexes. And, this is an apt description of the modern ETF market. Of course, things didn't always look this way. Prior to January 2008, the ETF market was rather boring. Every ETF during this period tracked some sort of pre-existing value-weighted market index. For a good example, think about the SPDR S&P 500 ETF, which tracks the S&P 500 and was the first ever ETF. But, in early 2008, the SEC changed its guidelines so that an ETF could now track its own self-defined benchmark. After this change, Invesco PowerShares was free to create an ETF that tracked the returns of the 100 lowest volatility S&P 500 stocks even if there was no pre-existing low-volatility S&P 500 index it could point to. All Invesco had to do was announce the weights involved in this low-volatility benchmark a day in advance.

The modern ETF industry closely resembles the environment that we model in Section

2. There are now more ETFs than stocks.[8] And, the sheer number and variety of these so-called *smart-beta* ETFs divides popular opinion.[9] "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."[10]

To be sure, niche ETFs tend to be smaller than broad value-weighted market ETFs, such as the well-known SPDR S&P 500 ETF. But, even the rebalancing activity of niche ETFs can affect a stock's underlying characteristics because ETFs execute the bulk of their trades during the final 20-to-30 minutes prior to market close.[11] 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."[12] A 2015 Goldman Sachs report found that energy-sector ETFs accounted for "15% of Chevron and 12% of Exxon Mobil's average daily volume" during a three-month stretch in 2015.[13]

Reason #2: Lack of Discretion. Second, the model applies to a setting where index-fund managers follow simple deterministic rebalancing rules. And, ETF managers don't deviate from their stated benchmarks like mutual-fund or hedge-fund managers do (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. 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 an ETF is underpriced. So, if arbitrageurs are constantly asking an ETF sponsor to create or redeem lots of shares, then the sponsor must be losing lots of money. As a result, creations and redemptions are only a small fraction of daily trading volume for ETFs, and these trades involve less than 0.5% percent of ETFs' net assets (Investment Company Institute, 2015).

Instead, ETF trading volume primarily comes from rebalancing activity just prior to

[8]Bloomberg. 5/16/2017. *Mutual Funds Ate the Stock Market. Now ETFs Are Doing It.*
[9]New York Times. 06/22/2017. *An Index-Fund Evangelist Is Straying From His Gospel.*
[10]Financial Times. 11/21/2017. *A ROSE by any other ticker symbol...*
[11]Wall Street Journal. 03/14/2018. *What's the Biggest Trade on the NYSE? The Last One.*
[12]Financial Times. 3/17/2017. *Machines and Markets: 5 Areas To Watch.*
[13]Goldman Sachs Equity Research. 04/10/2015. *ETFs: The Rise of the Machines.*

market close. This end-of-day trading is how ETFs make sure that there is very little difference between the market value of their end-of-day holdings and the market value of their stated benchmark. It's true that an ETF manager who does the bulk of his rebalancing activity 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 three-times more than stocks."[14] An investor interested in holding a smart-beta ETF is looking for quick access to a very targeted position. For a niche ETF, the additional trading costs incurred by end-of-day orders are nothing compared to the costs associated with replicating an entire position from scratch.

Reason #3: End-of-Day Data. Third and finally, the model relies on the fact that demand curves slope down (Shleifer, 1986). It's essential that the trading activity of one index fund can affect other index funds' benchmark composition. To observe these distortions empirically, we need to be able to see an index fund's holdings on a daily basis. And, we can observe end-of-day portfolio positions for ETFs. Note that other papers in the ETF literature, such as Ben-David et al. (2017), impute each ETF's daily portfolio position from its end-of-quarter financial statements. This is a perfectly reasonable approach for answering some research questions, but it won't work in our setting. We are interested in how the rebalancing decisions of different index funds interact with one another over the course of a few days. And, we can't study these interactions by imputing a fund's daily holdings from end-of-quarter reports.

Variable Construction. For each ETF, $f \in \{1, \ldots, F\}$, ETF Global provides data on assets under management, $\texttt{AUM}_{f,t}$, and relative portfolio weights in stock s, $\Omega_{f,s,t}$, at the end of each trading day t from January 1st, 2011 to December 31st, 2017. Thus, if $P_{s,t}$ is the price of stock s on day t, the actual number of shares of stock s that the fth ETF holds on day t, $Q_{f,s,t}$, is:

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

Total ETF trading volume for stock s on day t is given by $\texttt{ETFvlm}_{s,t} = \sum_{f=1}^{F} |Q_{f,s,t} - Q_{f,s,t-1}|$. We restrict our sample to only include ETFs that rebalance more than once a quarter. So, when viewing our results, you should have in mind the PowerShares S&P 500 Low-Volatility ETF rather than the SPDR S&P 500 ETF. We also exclude leveraged ETFs from our sample. We do this to emphasize that we are studying how rebalancing cascades can transmit an initial shock to stock A to an unrelated stock Z in an unpredictable way and not how leveraged ETFs predictably amplify initial shocks to stock A (Ivanov and Lenkey, 2018). We do not otherwise restrict, winsorize, or filter the data.

---

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

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

$$Q_{f,s,t} = \frac{1}{P_{s,t}} \times (\Omega_{f,s,t} \cdot \text{AUM}_{f,t})$$
$$= \frac{1}{P_{s,t}} \times (\Omega_{f,s,t-1} + \Delta\Omega_{f,s,t}^{rebal} + \Delta\Omega_{f,s,t}^{vw}) \cdot \text{AUM}_{f,t}$$

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

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

$$\Delta\Omega_{f,s,t}^{vw} = (R_{s,t}/R_{bmk,t} - 1) \cdot \Omega_{f,s,t-1} \tag{5}$$

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

So, to compute the amount of a stock's trading volume coming specifically from ETF rebalancing decision, we first calculate each ETF's predicted holdings for the stock on day t:

$$\bar{Q}_{f,s,t} = \frac{1}{P_{s,t}} \times (\Omega_{f,s,t-1} + \Delta\Omega_{f,s,t}^{vw}) \cdot \text{AUM}_{f,t}$$

Then, for each stock, we sum up the difference between every ETF's actual end-of-day holdings and this predicted value:

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

We use this as our daily measure of ETF rebalancing volume for each stock.

To measure the direction of ETF order flow (buy? or sell?), we also compute a measure of the order imbalance in each stock's ETF rebalancing volume:

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

This variable lies on the interval $[-1, 1]$. If $\texttt{ETFimbal}_{s,t} = -1$, then every share of stock s traded by ETFs for rebalancing reasons resulted in a sell order. Whereas, if $\texttt{ETFimbal}_{s,t} = +1$, then every single share of ETF rebalancing volume resulted in a buy order for stock s.

Summary Statistics. Table 1 provides summary statistics describing the ETF market. Panel A presents aggregate statistics at the monthly level. There were, on average, 1073.3 ETFs in our sample each month with a total of \$1.4 trillion in assets under management. These ETFs tracked almost 900 different benchmark indexes in the average month, and each ETF typically held positions in 263 different stocks. Panel B then presents cross-sectional statistics showing how these monthly aggregates varied across funds. While the average ETF in our sample had \$859.6 million in assets under management and held positions in 247 stocks, the median ETF only managed \$30.8 million and held positions in 78 stocks. These numbers confirm that the typical ETF in our sample is smaller than a large value-weighted market fund, such as the SPDR S&P 500 ETF [SPY]. But, as discussed earlier, even a relatively small ETF can have a large impact on the characteristics of one of its constituent stocks because ETFs do the bulk of their trading during the final 20-to-30 minutes of the trading day.

Next, Table 2 provides summary statistics describing the stocks that ETFs trade. Panel A describes ETF trading activity for each stock. The average stock in our sample is held by 476 different ETFs each month. The rebalancing activity of these ETFs typically resulted in an additional $e^{13.1} \approx 500k$ shares of each stock being traded each month, which represented $e^{13.1-15.5} \approx 10\%$ of total trading volume for the average stock in our sample. To emphasize the point that not all ETF trading volume is due to ETF rebalancing decisions, we also report total ETF trading volume. We find, however, that non-rebalancing volume makes up less than 1% of total ETF trading volume for the typical stock. Panel B then provides summary statistics describing more general characteristics of these stocks.

3.2 M&A Announcements

Having explained why we chose to study the ETF market, we now describe why we use M&A announcements for our set of initial shocks. We refer to the target of an M&A announcement as stock A. Our data on M&A deals comes from Thomson Financial. We use all deals that involve publicly traded target firms with an announcement date between January 1st, 2011 and December 31st, 2017. There are 1119 such deals in our seven-year sample period, yielding

an average of 14.3 announcements per month as shown in Table 3.

Effect on Rebalancing. 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 change immediately following the announcement (Andrade et al., 2001). And, while acquirers do not choose their M&A targets at random, the exact day that a deal is announced—Wednesday vs. Thursday—can be taken as exogenous. Consistent with these findings, Table 4 shows that ETFs rebalance their position in stock A on the day that stock A gets revealed as the target of an M&A deal. Let t_A denote the day stock A is announced as an M&A target. We create a panel dataset containing the ETF rebalancing volume for each stock A during the 26-day window $t \in \{t_A - 20, \ldots, t_A + 5\}$. Then, we regress stock A's ETF rebalancing volume on indicator variables for the number of days until the M&A announcement:

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

In the equation above, α_A and α_{mmyy} denote stock-A and month-year fixed effects respectively.

The first column in Table 4 shows that ETF rebalancing volume for stock A rises by 139.69% on the exact day that stock A is announced as an M&A target. The second column then shows results for the same regression specification after including controls for stock A's lagged total trading volume. The point estimate for the effect of an M&A announcement on ETF rebalancing volume in stock A hardly changes when moving from the first to the second column. The third column shows the results of a similar specification that also includes additional indicator variables for the days (t_A-2), (t_A-3), (t_A-4), and (t_A-5) prior to stock A's announcement. This column reveals that there is no pre-trend in ETF managers' reaction to the M&A announcement. This timing is consistent with the fact that ETF managers don't have any discretion when it comes to deviating from their benchmark index overnight.

Placebo Test. The fifth and final column of Table 4 shows the results of the same regression specification using placebo announcement dates for each stock A. We randomly re-assign the announcement date t_A to some other point in our sample period prior to each stock A's actual announcement. Consistent with the idea that it's the M&A announcement itself that's causing the jump in ETF rebalancing volume, we find that there is no jump in ETF rebalancing volume on these placebo dates. And, we find that the coefficients on each stock A's total trading volume remain unchanged, which suggests our result isn't being driven by some broader trading-volume anomaly that occurs around the time of each M&A announcement.

3.3 Experimental Design

The model predicts that, following an initial shock to stock A, i) stock Zs with more neighbors in the ETF rebalancing network should be involved in ETF rebalancing cascades more often,

but ii) it shouldn't be possible to predict the direction of these cascades' effect on a stock Z's demand. Here's how we use our data on ETF rebalancing activity in the wake of M&A announcements to test these two predictions.

Unrelated Stock Zs. To start with, we need to make sure that any ETF rebalancing activity we measure is due to ETF rebalancing cascades and not some omitted variable affecting both stock A and stock Z. So, we create a separate panel for each M&A announcement containing the set of stock Zs that are unrelated to stock A during the 26-day window $t \in \{t_A - 20, \ldots, t_A - 1, t_A, t_A + 1, \ldots, t_A + 5\}$ around stock A's announcement. For a stock Z to be unrelated to stock A, the target of an M&A announcement, these two stocks have to be twice removed in the network of ETF rebalancing decisions. Stock Z can't have been rebalanced at any point during the last month by any ETF that also held stock A during the last month. And, if stock B and stock A were both held at any point during the last month by the same ETF, then stock Z can't have been rebalanced during the last month by any ETF that also held stock B during that time period. In other words, the chain of ETF rebalancing decisions from stock A to stock Z has to be $A \rightarrow B \rightarrow C \rightarrow Z$ or longer. Because there are smart-beta ETFs tracking things like large-cap, value, and industry, this twice-separated criteria implies that stock A and stock Z don't have any similar factor exposures and don't share any well-known firm characteristics. We then combine these separate datasets—one for each of the 1119 M&A announcements in our sample—into a single panel dataset indexed by M&A announcement, stock Z, and date. Because the same stock Z can be affected by ETF rebalancing cascades starting with different initial stock As, we index the rows of this dataset with the subscript $Z,t|A$.

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

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

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

The model then makes predictions about the differential effect of an index-fund rebalancing cascade on stock Zs depending on their number of neighbors in the ETF rebalancing network. So, the second difference will capture whether or not a particular stock Z has lots of neighbors in the ETF rebalancing network. We say that stock s is a neighbor to stock Z if an ETF

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

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

We calculate the number of neighbors for each stock Z using data from the month prior to stock A's M&A announcement. So, this indicator variable does not vary over the 26-day window surrounding each target stock's M&A announcement, which is why we write $\texttt{manyNhbr}_{Z|A}$ rather than $\texttt{manyNhbr}_{Z,t|A}$. However, because we calculate the median number of stock-Z neighbors separately for each M&A announcement, this indicator variable can vary across M&A announcements. The same stock Z can have an above-median number of neighbors in the ETF rebalancing network relative to one M&A announcement but a below-median number of neighbors relative to another. This is why we write $\texttt{manyNhbr}_{Z|A}$ rather than just $\texttt{manyNhbr}_Z$.

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

$$\begin{aligned} \texttt{ETFrebal}_{Z,t|A} = {} & \alpha_{\texttt{\#nhbr}|A} + \alpha_Z + \beta \cdot \texttt{afterAncmt}_{Z|A} \\ & + \gamma \cdot \left\{ \texttt{afterAncmt}_{t|A} \times \texttt{manyNhbr}_{Z|A} \right\} \qquad (6) \\ & + \varepsilon_{Z,t|A} \end{aligned}$$

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

fact that ETFs might always be more likely to trade some stocks than others, regardless of how many neighbors they have.

Next, let's consider the coefficients on the first difference, β. This coefficient captures the rise in ETF rebalancing volume for all stock Zs in the five days immediately after an M&A announcement about an unrelated stock A. Because an ETF rebalancing cascade has the potential to affect the demand for all stocks—i.e., even stock Zs with few neighbors—we should expect the average ETF rebalancing volume of all stock Zs to rise in five days after the M&A announcement. In other words, we should expect to estimate $\beta > 0$ in Equation (6). After all, there are more benchmark indexes than stocks in modern financial markets. This means that most stocks in the market will be connected to one another by at least one direct neighbor (Erdos and Rényi, 1960), and it's possible for these stocks Zs to be affected by an ETF rebalancing cascade. The key prediction of the model is that it's more likely for stock Zs with many neighbors to be hit.

The coefficient γ then captures how much more the ETF rebalancing volume increases for stock Zs with many neighbors than for stock Zs with few neighbors in the five days immediately after stock A's M&A announcement. The first key prediction of the model is that we should estimate $\gamma > 0$ in Equation (6). By contrast, Proposition 2.3b suggests that it shouldn't be possible to predict the direction (buy? or sell?) of the resulting demand shocks. So, the second key prediction of the model is that, if we replace ETF rebalancing volume with ETF order imbalance on the left-hand side of Equation (6),

$$
\begin{aligned}
\mathtt{ETFimbal}_{Z,t|A} = {} & \alpha_{\mathtt{\#nhbr}|A} + \alpha_Z + \beta \cdot \mathtt{afterAncmt}_{Z|A} \\
& + \gamma \cdot \{\mathtt{afterAncmt}_{t|A} \times \mathtt{manyNhbr}_{Z|A}\} \qquad (7) \\
& + \varepsilon_{Z,t|A},
\end{aligned}
$$

then we should estimate $\gamma = 0$. We don't include the level effect for $\mathtt{manyNhbr}_{Z|A}$ in our regression specification because it gets subsumed by the announcement$\times$#nhbr fixed effects.

Source of Identification. At this point, it's important to pause and spell out the source of our identification in these regressions. We want to emphasize that we're not making an assumption that stock Zs with many neighbors are similar stock Zs with few neighbors in the ETF rebalancing network. In fact, there are good reasons to expect these stocks to be different. M&A announcements are just one kind of initial shock that might trigger an ETF rebalancing cascade. So, if you really believe that ETF rebalancing cascades generate demand noise, then you should expect stock Zs with an above-median number of neighbors to always have more ETF rebalancing volume than below-median stock Zs since these stocks will be hit by more ETF rebalancing cascades. And, the summary statistics in Table 5 bear this out.

Instead, our identification is coming from the timing of the M&A announcements. We're
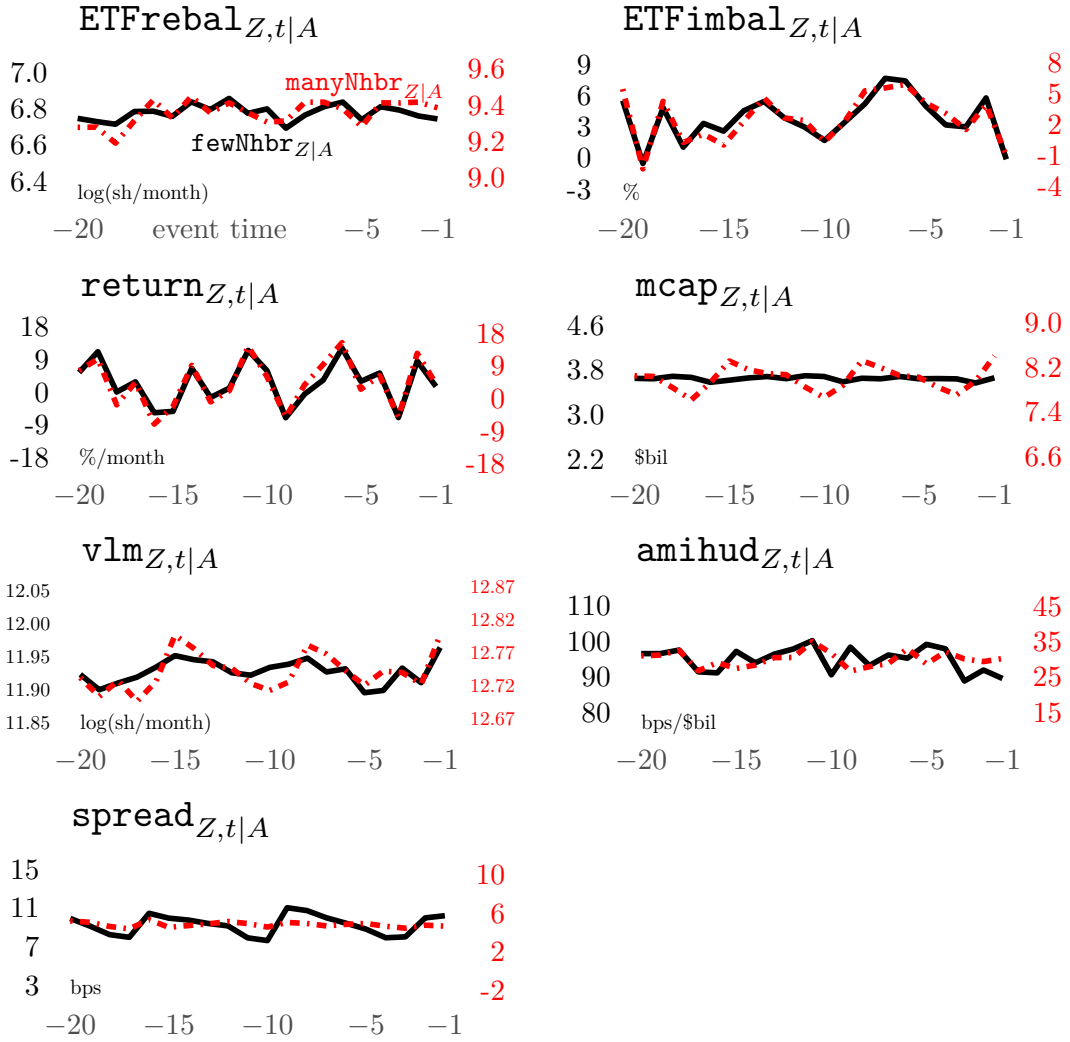
Figure 5. No Pre-Trends. Average ETF activity and characteristics of stock Zs during the 20 days prior to an M&A announcement about stock A. x-axis: event time with M&A announcement occurring on day 0. Red, Dashed: average value for stock Zs with an above-median number of neighbors; right y-axis. Black, Solid: average for stock Zs with a below-median number of neighbors; left y-axis. $\texttt{ETFrebal}_{Z,t|A}$: ETF rebalancing volume reported on a base-e logarithmic scale. $\texttt{ETFimbal}_{Z,t|A}$: ratio of signed ETF rebalancing volume to total ETF rebalancing volume in percent. $\texttt{return}_{Z,t|A}$: return in percent per month. $\texttt{mcap}_{Z,t|A}$: market capitalization in billions of dollars. $\texttt{vlm}_{Z,t|A}$: number of shares traded per month reported on a base-e logarithmic scale. $\texttt{amihud}_{Z,t|A}$: Amihud (2002) illiquidity measure over previous 20 days in basis points per \$1 million order. $\texttt{spread}_{Z,t|A}$: bid-ask spread as a fraction of the daily midpoint in basis points.

ORCL   ORCL   ORCL   INCB

MAI   NET

Jul. 21, 2014   Jun. 19, 2012   Jan. 25, 2012
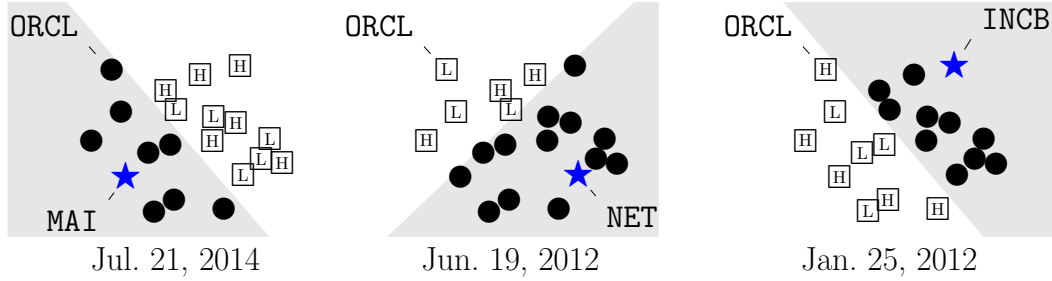
Figure 6. Time-Varying Comparison Groups. Each panel depicts the the same set of stocks during 3 different M&A announcements: Owens & Minor's purchase of Medical Action Industries [MAI] announced on Jul. 21, 2014; Sonus Networks' purchase of Network Equip Technologies [NET] announced on Jun. 19, 2012; and, Old National Bancorp's purchase of Indiana Community Bancorp [INCB] announced on Jan. 25, 2012. The target of each M&A announcement, stock A, is denoted by a blue star. Each black circle denotes a stock that'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.

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

This identification strategy raises two kinds of concerns. First, you might be worried that something else about the set of unrelated stock Zs is changing at the time of the M&A announcements. This is something we can test in the data. Figure 5 shows that, although above- and below-median stock Zs tend to have different amounts of ETF trading activity, this difference is constant in the run up to each M&A announcement. The figure also shows that the same statement holds for other stock Z characteristics, such as realized returns, market capitalization, trading volume, and liquidity. Again, we should expect to observe differences between above- and below-median stock Zs since above-median stock Zs are always more likely to be hit by ETF rebalancing cascades. It's the fact that these differences

suddenly change in the wake of an M&A announcement about an unrelated stock which provides evidence of ETF rebalancing cascades.

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

What's more, because we are including stock-Z fixed effects in Equations (6) and (7), any stock Z that always has either an above- or below-median number of neighbors in the ETF rebalancing network will not contribute to our coefficient estimates. For example, the ETF rebalancing volume of Apple, Inc. is not reflected in our analysis because Apple always has an above-median number of neighbors in the ETF rebalancing network. Thus, our findings can't be explained by ETFs always trading some stock Zs differently than others. Any omitted variable would have to account for why ETFs suddenly change their rebalancing behavior for only the stock Zs that have an above-median number of neighbors relative to a particular stock A in the five days immediately after that stock A's M&A announcement.

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

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

3.4 Estimation Results

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

ETF Rebalancing Volume. Table 6 describes how cascades affect ETF rebalancing volume by reporting the estimated coefficients for the regression specification in Equation (6). The first column shows that ETF rebalancing volume for all stock Zs tends to rise by $\beta = 2.64\%$ on average in the five days immediately after an M&A announcement about an unrelated stock A. But, the third column shows that this growth is concentrated among stock Zs that have many neighboring stocks in the network defined by ETF rebalancing rules. Consistent with the model, we find that ETF rebalancing volume is $\gamma = 2.06\%$ higher for above-median stock Zs than for below-median stock Zs in the five days immediately following an M&A announcement about an unrelated stock A. The second and fourth columns of Table 6 confirm that the sudden spike in the ETF rebalancing volume isn't due to a general run-up in trading volume. When we include stock Z's total trading volume in our regression specification, our point estimate for γ remains largely unchanged.

We run two different kinds of placebo tests to make sure that our estimate of $\gamma \approx 2\%$ is due to ETF rebalancing cascades and not some omitted variable. The results of the first placebo test can be found in the fifth column of Table 6. For this column, we re-estimated the regression specification in Equation (6) using data during the 26-day window surrounding $t_A - 30$ rather than t_A. If we shift stock A's announcement date forward by 30 days, then we can be certain that there was no initial M&A announcement and thus no subsequent ETF rebalancing cascade. And, when we do this, our point estimate for γ shrinks by a factor of ten, from 2.06% to 0.20%, and becomes statistically indistinguishable from zero. What's more, this lack of statistical significance is being completely driven by the smaller coefficient estimate. The lack of statistical significance isn't due to a lack of power. Our standard error on γ is roughly the same, 0.35 vs. 0.32, in both the true and the placebo samples.

We report the results of the second placebo test in Table 7. The idea behind this test is a little different. Before we checked to make sure that there was no difference between

the ETF rebalancing volume of above- and below-median stock Zs following alternative announcement dates when no initial shock to stock A took place. Now, we're going to use the right announcement date and instead restrict our attention to the rebalancing volume coming from the 30% of ETFs in our sample that are the least likely to rebalance each day. All of the ETFs in our sample rebalance more than once a quarter. But, some of the ETFs rebalance much more than others—once a day vs. twice a quarter. And, the ETFs that rebalance the least should also be the least likely to transmit the effects of an ETF rebalancing cascade. Consistent with this prediction, when we restrict our attention to these infrequent rebalancers, above-median stock Zs have ETF rebalancing volume that's statistically indistinguishable from below-median stock Zs in the five days after an M&A announcement. These two placebo tests provide strong evidence that $\gamma \approx 2\%$ is due to ETF rebalancing cascades.

ETF Order Imbalance. Next, in Table 8, we give evidence that the extra ETF rebalancing volume experienced by above-median stock Zs following an unrelated M&A announcement is no more likely to be made up of buy orders than of sell orders. This table reports the estimated coefficients for the regression specification in Equation (7). The point estimate of $\gamma = 0.74\%$ with a standard error of 0.51% in second column reveals that there's no statistically measurable difference between the ETF order imbalance of above- and below-median stock Zs following an M&A announcement. Taken together, this evidence suggests that, while it'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.

Price Impact. Although ETF rebalancing demand is no more likely to be positive than negative on average, Table 9 reveals that it still has a significant effect on prices. The first two columns in this table show the results of replacing the dependent variable in Equation (6) with the daily returns of each stock Z. On average, the prices of above-median stock Zs look just like the prices of below-median stock Zs following an initial shock to an unrelated stock A. The third and fourth columns of Table 9 then show the differential effect of very positive or very negative demand from ETF rebalancing on the prices of these same subgroups. Because above-median stock Zs tend to realize larger demand shocks, their prices rise by 14bps per day more than those of below-median stock Zs in response to order flow composed of a larger fraction of buy orders. And, their prices shrink by an extra 21bps per day in response of order flow with a higher proportion of sell orders.

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

to get a better sense of the total amount of demand noise produced by ETF rebalancing cascades irrespective of the initial shock. We regress each stock's log ETF rebalancing volume normalized by its standard deviation over the past twelve months, $\hat{\sigma}_{s,t}$, on the number of neighbors this stock has in the ETF rebalancing network in thousands:

$$\texttt{ETFrebal}_{s,t}/\hat{\sigma}_{s,t} = \alpha + \beta \cdot \texttt{\#nbrs}_{s,t} + \varepsilon_{s,t} \tag{8}$$

The first column of Table 10 reports that $\beta = 0.40$, which implies that a 1σ increase in the number of neighbors a stock has is associated with a 0.4σ increase in a stock's log ETF rebalancing volume. Moreover, the second column of Table 10 shows that this point estimate remains statistically significant even if we include fixed effects for the number of ETFs that hold the stock, suggesting that the effect is operating through a stock's position in the ETF rebalancing network and not merely through the number of ETFs that directly hold the stock.

4   Conclusion

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

This paper proposes an analogous explanation for where seemingly random demand shocks come from in financial markets. A stock's demand might appear random, not because individual investors are actually 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. 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 ETFs.

A natural next question is: *'What should a researcher or trader do with this information about where demand noise comes from?'* First, it's useful for researchers to know that complexity generates noise in financial markets because this mechanism makes predictions about where demand noise will be loudest. This insight offers news ways to test existing asset-pricing models. Consider your favorite limits-to-arbitrage model. In the past, if you wanted to test this model, then you would just look for situations where arbitrageurs were most constrained. But, this is only half of the battle. In order for the limits of arbitrage to bind, there must also be a non-fundamental shock (Chinco, 2018). And, our noise-generating mechanism says that stocks with more neighbors in the ETF rebalancing network will realize more of these

shocks. So, you can now check whether the implications of the limits-to-arbitrage model are strongest for the constrained stocks with the most noise.

The practical implication is a bit different for traders. The way to use the insights in this paper 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. Market participants can see whether each unrelated stock Z has many or few neighbors. And, because the stock Zs with many neighbors are more likely to be hit by an ETF rebalancing cascade, market participants will realize that they are more likely to see erratic non-fundamental demand shocks for these stocks. So, the stock Zs with many neighbors should have higher liquidity. And, the remaining columns in Table 10 bear out this prediction. Our results suggest that you could implement any long-short strategy more efficiently by focusing each leg on the stocks with the most neighbors in the ETF rebalancing network.

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.*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Brunnermeier, M. and L. H. Pedersen (2008). Market liquidity and funding liquidity. *Review of Financial Studies*.

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

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

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

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

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

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

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

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*.

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

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*.

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

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

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

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*.

Loughran, T. and B. McDonald (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *Journal of Finance*.

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*.

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

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

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

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

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

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

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

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 $\{0,1\}^\star = \cup_{n=0,1,2,\dots}\{0,1\}^n$ denote the set of binary strings.

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

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

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 $F : \{0,1\}^\star \mapsto \{0,1\}^\star$ such that

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

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 (2) exists with or without the path described by Equation (3). This means that if $(Z, \mathsf{M}, T, \{s\}) \in$ If and $(Z, \mathsf{M}, T, \{s'\}) \in$ If, then $(Z, \mathsf{M}, T, \{s, s'\}) \in$ 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). $\qquad\square$

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 $O[S^k]$ steps for some $k > 0$. Whereas, Prob_2 requires exponential time if every solution algorithm requires $2^{\ell \cdot S}$ steps on at least one instance for some $\ell > 0$.

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

"If cases like this regularly arose in practice, then it would'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 4)**.** 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 κ/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

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

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

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

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

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

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

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

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

The second line implies that $\mathrm{E}[\hat{\mathsf{S}}_{\max}(Z, \mathsf{M}, T)]$ 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. □

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 (3Sat).
- *Instance: A Boolean formula defined over N input variables*

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

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

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

Corollary. Let Prob denote any decision problem. If Prob is reducible to 3Sat, then Prob is NP complete.

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

STEP 1: First, create variables to track of the state of the rebalancing cascade:
- For each possible value of $(x_{s,t}, \Delta x_{s,t})$,

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

define 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:

$$\left(\overline{\alpha(0,0)}_{s,t} \vee \overline{\alpha(1,1)}_{s,t}\right)$$

- Second, there are constraints that encode the rebalancing cascade updating rules. e.g., if stock s has one negative neighbor, s', and one positive neighbor, s'', then the rebalancing-cascade rules are encoded in four different clauses:

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

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

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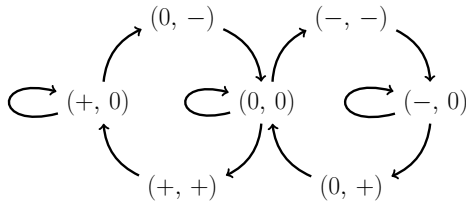
Figure 7. State Diagram. All possible ways that a single stock could move between the 7 possible values of $(x_{s,t}, \Delta x_{s,t})$ in successive rounds of an index-fund rebalancing cascade. Arrows denote transitions. Loops denote unchanged values in successive rounds.

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

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

Problem C (Majority).
- *Instance: A Boolean formula defined over N input variables*

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

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

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

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

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

Problem D (2Sat).
- *Instance: A Boolean formula defined over N input variables*

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

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

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

Proof (Proposition 2.4a). 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 $s \in$ 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* (1) *was changed to the following for some $\theta \in (0, 1)$:*

$$u_{s,t+1} = \frac{1}{|\mathsf{In}^+_{s,t+1}| + |\mathsf{In}^-_{s,t+1}|} \cdot \left(\sum_{s' \in \mathsf{In}^-_{s,t+1}} x_{s',t} - \sum_{s'' \in \mathsf{In}^+_{s,t+1}} x_{s'',t} \right)$$

$$x_{s,t+1} = \theta \cdot (x_{s,t} + u_{s,t+1})$$

- *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* $\mathsf{SmoothHow}$. *There's a polynomial-time algorithm, F, such that for any $\delta > 0$*

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

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 $\mathsf{SmoothHow}$, 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'. □

Proof (Equation 5). Let $R_{s,t}$ denote the day-t return on stock s, and let $R_{b,t}$ denote the day-t return on the fth ETF's value-weighted benchmark index. Similarly, let $size_{s,t}$ denote the market capitalization of stock s on day t, and let $size_{b,t}$ denote the market capitalization of the fth ETF's benchmark index on day t. Then, the change in the fth ETF's value-weighted portfolio weight for stock s, $\Omega_{f,s,t}^{vw} = size_{s,t}/size_{b,t}$, is given by:

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

□

B Tables

Summary Statistics, ETF Market

Panel A. Time Series

	Avg	Sd	25%	50%	75%
count_t	1073.3	168.2	914	1017	1249
$\text{\#benchmark}_t$	896.9	141.7	769	836	1045
$\text{\#stock}_t$	263.4	15.6	256.7	269.5	274.6
AUM_t	1390.0	651.8	1060.9	1485.4	1701.0

Panel B. Cross-Section

	Avg	Sd	25%	50%	75%
$\text{\#stock}_{f,t}$	247.0	499.2	30	78	254
$\text{AUM}_{f,t}$	859.6	5381.3	4.8	30.8	217.6

Table 1. Summary statistics for the ETF market using data from ETF Global. Sample: January 2011 to December 2017. **Panel A.** Aggregate statistics for the entire ETF market each month. count_t: number of ETFs in sample. $\text{\#benchmark}_t$: number of different benchmarks used by these ETFs. $\text{\#stock}_t$: average number of stocks held by each ETF in a given month. AUM_t: total assets under management for all ETFs in billions of dollars. **Panel B.** Cross-sectional statistics of the ETF market computed using fund×month observations. $\text{\#stock}_{f,t}$: number of stocks held by an ETF. $\text{AUM}_{f,t}$: assets under management in millions of dollars.

Summary Statistics, All Stocks

Panel A. ETF Activity

	Avg	Sd	25%	50%	75%
#ETFholding$_{s,t}$	476.1	453.7	88	366	713
#nhbr$_{s,t}$	3.4	1.1	3.3	3.8	4.0
ETFvlm$_{s,t}$	13.2	3.4	11.7	13.4	14.7
ETFrebal$_{s,t}$	13.1	3.5	11.5	13.2	14.6
ETFimbal$_{s,t}$	−1.2	35.6	−19.9	0.0	19.4

Panel B. Characteristics

	Avg	Sd	25%	50%	75%
return$_{s,t}$	1.1	11.5	−4.2	1.1	6.4
mcap$_{s,t}$	5.9	23.0	0.2	0.8	3.2
vlm$_{s,t}$	15.5	2.0	14.3	15.7	16.9
amihud$_{s,t}$	45.6	253.8	0.1	0.2	2.8
spread$_{s,t}$	46.3	89.3	4.2	10.2	38.1

Table 2. Summary statistics using stock×month level observations. Sample: January 2011 to December 2017. **Panel A.** ETF activity for each stock. #ETFholding$_{s,t}$: number of ETFs that hold a stock. #nhbr$_{s,t}$: number of neighboring stocks in the ETF rebalancing network in thousands. ETFvlm$_{s,t}$: ETF trading volume each month on a base-e logarithmic scale. ETFrebal$_{s,t}$: ETF rebalancing volume each month on a base-e logarithmic scale. ETFimbal$_{s,t}$: signed ETF rebalancing volume divided by total ETF rebalancing volume reported in percent. **Panel B.** Characteristics of each stock. return$_{s,t}$: return in percent per month. mcap$_{s,t}$: market capitalization in billions of dollars. vlm$_{s,t}$: total trading volume each month on a base-e logarithmic scale. amihud$_{s,t}$: Amihud (2002) illiquidity measure computed daily within a given month in basis points per $1 million order. spread$_{s,t}$: average bid-ask spread as a fraction of midpoint during month in basis points.

M&A Announcements

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

Table 3. Number of M&A announcements about a publicly traded target firm each month. Data comes from Thompson Financial. Sample: January 2011 to December 2017. There are 1199 total M&A announcements in our sample. $\texttt{\#ancmt}_t$: number of M&A announcements per month. $\texttt{\#stockZ}_t$: number of unrelated stock Zs for M&A announcements in a given month in thousands.

ETF Rebalancing Volume, Stock A

Dependent Variable: `ETFrebal`$_{A,t}$

	Actual Announcements				Placebo
	(1)	(2)	(3)	(4)	(5)
$1_{\{t=t_A+1\}}$	76.24*** (9.35)	27.27*** (9.71)	76.24*** (9.37)	27.27*** (9.73)	−9.30 (11.88)
$1_{\{t=t_A\}}$	139.69*** (8.09)	137.35*** (8.14)	139.70*** (8.11)	137.36*** (8.16)	9.75 (11.91)
$1_{\{t=t_A-1\}}$	2.79 (9.64)	1.27 (9.62)	2.80 (9.65)	1.28 (9.63)	21.27 (13.82)
$1_{\{t=t_A-2\}}$			9.24 (10.11)	9.51 (9.98)	−8.31 (17.62)
$1_{\{t=t_A-3\}}$			8.34 (10.80)	7.98 (10.75)	0.40 (13.48)
$1_{\{t=t_A-4\}}$			0.81 (11.39)	1.13 (11.31)	8.06 (11.86)
$1_{\{t=t_A-5\}}$			−11.79 (9.77)	−10.16 (9.73)	8.08 (14.29)
`vlm`$_{A,t-1}$		17.28*** (0.42)		17.28*** (0.42)	17.30*** (0.42)
`vlm`$_{A,t-2}$		5.56*** (0.24)		5.56*** (0.24)	5.55*** (0.24)
`vlm`$_{A,t-3}$		2.92*** (0.28)		2.92*** (0.28)	2.92*** (0.28)
Month-Year FE	Y	Y	Y	Y	Y
Stock-A FE	Y	Y	Y	Y	Y
R^2	67.0%	67.4%	67.0%	67.4%	67.4%
Observations	7,206,116	7,187,021	7,206,116	7,187,021	7,187,021

Table 4. Effect of initial M&A announcement about stock A on ETF rebalancing volume for the same stock A. Sample: January 1st, 2011 to December 31st, 2017. Each column represents results from a separate regression using daily data on the 26-day window surrounding each M&A announcement, $\{t_A - 20, \ldots, t_A - 1, t_A, t_A + 1, \ldots, t_A + 5\}$. Columns (1)-(4) report results for actual announcements. Column (5) reports results using a randomly assigned announcement date for each M&A target in our sample. `ETFrebal`$_{A,t}$: dependent variable is the ETF rebalancing volume for stock A on date t reported on a base-e logarithmic scale. Coefficient estimate of 1 indicates a 1% increase in a stock's daily ETF rebalancing volume. $1_{\{t=t_A-h\}}$: indicator variable that is one if an observation was made h days prior to stock A's announcement date. `vlm`$_{A,t}$: total trading volume for stock A on day t reported on a base-e logarithmic scale. Numbers in parentheses are standard errors clustered by stock. Statistical significance: $\star = 10\%$, $\star\star = 5\%$, and $\star\star\star = 1\%$. *Reads: "Stock A has* 139.69% *more ETF rebalancing volume on the day it is announced as an M&A target."*

Summary Statistics, Stock Z

Panel A. ETF Activity

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

Panel B. Characteristics

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

Table 5. Summary statistics for the set of stock Zs unrelated to each M&A announcement. An observation is the average value for a stock Z during the 20 days leading up at an M&A announcement. Data on M&A announcements with publicly traded target firms comes from Thompson Financial. Stock-market data comes from CRSP. Sample: January 1st, 2011 to December 31st, 2017. All: all stock Zs that are unrelated to each M&A announcement. Many Neighbors: stock Zs with an above-median number of neighbors for a particular M&A announcement. Few Neighbors: stock Zs with a below-median number of neighbors for a particular M&A announcement. **Panel A.** ETF activity for each stock. #ETFholding$_{Z|A}$: number of ETFs that hold a stock. #nhbr$_{Z|A}$: number of neighboring stocks in the ETF rebalancing network in thousands. ETFvlm$_{Z|A}$: ETF trading volume each month on a base-e logarithmic scale. ETFrebal$_{Z|A}$: ETF rebalancing volume each month on a base-e logarithmic scale. ETFimbal$_{Z|A}$: signed ETF rebalancing volume divided by total ETF rebalancing volume reported in percent. **Panel B.** Characteristics of each stock. return$_{Z|A}$: return in percent per month. mcap$_{Z|A}$: market capitalization in billions of dollars. vlm$_{Z|A}$: total trading volume each month on a base-e logarithmic scale. amihud$_{Z|A}$: Amihud (2002) illiquidity measure computed daily within a given month in basis points per $1 million order. spread$_{Z|A}$: average bid-ask spread as a fraction of midpoint during month in basis points.

ETF Rebalancing Volume, Stock Z

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

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

Table 6. Effect of an initial M&A announcement about stock A on the ETF rebalancing volume for an unrelated stock Z. Sample: January 1st, 2011 to December 31st, 2017. Each column presents the results of a separate regression using daily data on the 26-day window surrounding each M&A announcement, $\{t_A - 20, \ldots, t_A - 1, t_A, t_A + 1, \ldots, t_A + 5\}$. Columns (1)-(4) report results for actual announcements. Column (5) reports results using a placebo sample where each M&A announcement is re-assigned to date $t_A - 30$. $\texttt{ETFrebal}_{Z,t|A}$: dependent variable is ETF rebalancing volume for stock Z on date t following M&A announcement about stock A on a base-e logarithmic scale; coefficient of $+1$ indicates a 1% per day increase in a stock's ETF rebalancing volume. $\texttt{afterAncmt}_{t|A}$: indicator variable that is one during the five days following an M&A announcement about stock A. $\texttt{manyNhbr}_{Z|A}$: indicator variable that is one if stock Z has an above-median number of neighbors in the ETF rebalancing network relative to the M&A announcement about stock A. $\texttt{vlm}_{Z,t|A}$: total trading volume for stock Z on a given day reported on a base-e logarithmic scale. Numbers in parentheses are standard errors clustered by stock Z. Statistical significance: $\star$ = 10%, $\star\star$ = 5%, and $\star\star\star$ = 1%. *Reads: "In the five days after an M&A announcement about stock A, unrelated stock Zs with an above-median number of neighbors in the ETF rebalancing network realize 2.06% per day more ETF rebalancing volume than unrelated stock Zs with a below-median number of neighbors."*

ETF Rebalancing Volume, Stock Z
Least-Active ETFs

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

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

Table 7. Effect of initial M&A announcement about stock A on the ETF rebalancing volume of an unrelated stock Z when looking only at ETF rebalancing volume for the 30% of ETFs that rebalance their positions the least frequently. Sample: January 1st, 2011 to December 31st, 2017. Each column presents the results of a separate regression using daily data on the 26-day window surrounding each M&A announcement, $\{t_A - 20, \ldots, t_A - 1, t_A, t_A + 1, \ldots, t_A + 5\}$. $\texttt{ETFrebal}_{Z,t|A}$: dependent variable is ETF rebalancing volume for stock Z on date t following M&A announcement about stock A on a base-e logarithmic scale; coefficient of $+1$ indicates a 1% per day increase in a stock's ETF rebalancing volume. $\texttt{afterAncmt}_{t|A}$: indicator variable that is one during the five days following an M&A announcement about stock A. $\texttt{manyNhbr}_{Z|A}$: indicator variable that is one if stock Z has an above-median number of neighbors in the ETF rebalancing network relative to the M&A announcement about stock A. $\texttt{vlm}_{Z,t|A}$: total trading volume for stock Z on a given day reported on a base-e logarithmic scale. Numbers in parentheses are standard errors clustered by stock Z. Statistical significance: $\star = 10\%$, $\star\star = 5\%$, and $\star\star\star = 1\%$. *Reads: "When we restrict our attention to the 30% of ETFs that are the least likely to rebalance each day, stock Zs with many neighbors have ETF rebalancing volume that's statistically indistinguishable from stock Zs with few neighbors in the five days after an M&A announcement."*

ETF Order Imbalance, Stock Z

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

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

Table 8. Effect of initial M&A announcement about stock A on the ETF order imbalance for an unrelated stock Z. Sample: January 1st, 2011 to December 31st, 2017. Each column presents the results of a separate regression using daily data on the 26-day window surrounding each M&A announcement, $\{t_A - 20, \ldots, t_A - 1, t_A, t_A + 1, \ldots, t_A + 5\}$. $\texttt{ETFimbal}_{Z,t|A}$: dependent variable is ETF order imbalance for stock Z on date t following M&A announcement about stock A; coefficient estimate of $+1$ indicates a 1% per day increase in a stock's ETF order imbalance. $\texttt{afterAncmt}_{t|A}$: indicator variable that is one during the five days following an M&A announcement about stock A. $\texttt{manyNhbr}_{Z|A}$: indicator variable that is one if stock Z has an above-median number of neighbors in the ETF rebalancing network relative to the M&A announcement about stock A. $\texttt{vlm}_{Z,t|A}$: total trading volume for stock Z on a given day reported on a base-e logarithmic scale. Numbers in parentheses are standard errors clustered by stock Z. Statistical significance: $^\star = 10\%$, $^{\star\star} = 5\%$, and $^{\star\star\star} = 1\%$. *Reads: "Although ETF rebalancing volume is higher for stocks Zs with many neighbors than for stock Zs with few neighbors in the five days after an M&A announcement about stock A, this ETF rebalancing volume is no more likely to consist of buy orders than of sell orders."*

Price Impact, Stock Z

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

	(1)	(2)	(3)	(4)
$\texttt{posETFimbal}_{Z,t|A} \times \texttt{afterAncmt}_{t|A} \times \texttt{manyNhbr}_{Z|A}$			14.17*** (1.21)	14.29*** (1.21)
$\texttt{afterAncmt}_{t|A} \times \texttt{manyNhbr}_{Z|A}$	−0.45 (0.32)	−0.47 (0.32)	1.13*** (0.42)	1.07*** (0.42)
$\texttt{negETFimbal}_{Z,t|A} \times \texttt{afterAncmt}_{t|A} \times \texttt{manyNhbr}_{Z|A}$			−21.12*** (1.29)	−20.92*** (1.28)
$\texttt{afterAncmt}_{t|A}$	−1.05*** (0.23)	−1.16*** (0.23)	−2.66*** (0.30)	−2.63*** (0.30)
$\texttt{posETFimbal}_{Z,t|A}$			0.71* (0.40)	0.23 (0.40)
$\texttt{negETFimbal}_{Z,t|A}$			−3.19*** (0.39)	−3.49*** (0.39)
$\texttt{vlm}_{Z,t|A}$		8.36*** (0.34)		8.27*** (0.34)
Announcement×#nhbr FE	Y	Y	Y	Y
Stock-Z FE	Y	Y	Y	Y
Additional Interactions	N	N	Y	Y
R^2	2.2%	2.3%	2.6%	2.6%
Observations		23,632,529		

Table 9. Effect of ETF order imbalance on stock Z's returns in the days after an M&A announcement about an unrelated stock A. Sample: January 1st, 2011 to December 31st, 2017. Each column presents the results of a separate regression using daily data on the 26-day window surrounding each M&A announcement, $\{t_A - 20, \ldots, t_A - 1, t_A, t_A + 1, \ldots, t_A + 5\}$. $\texttt{return}_{Z,t|A}$: dependent variable is the return of stock Z on date t; coefficient of +1 indicates a 1bps increase in a stock's daily return. $\texttt{afterAncmt}_{t|A}$: indicator variable that is one during the five days following an M&A announcement about stock A. $\texttt{manyNhbr}_{Z|A}$: indicator variable that is one if stock Z has an above-median number of neighbors in the ETF rebalancing network relative to the M&A announcement about stock A. $\texttt{posETFimbal}_{Z,t|A}$: indicator variable that is one if stock Z has an above-75%tile ETF order imbalance on day t. $\texttt{negETFimbal}_{Z,t|A}$: indicator variable that is one if stock Z has a below-25%tile ETF order imbalance on day t. $\texttt{vlm}_{Z,t|A}$: total trading volume for stock Z each day reported on a base-e logarithmic scale. Numbers in parentheses are standard errors clustered by stock Z. Statistical significance: $\star = 10\%$, $\star\star = 5\%$, and $\star\star\star = 1\%$. *Reads: "While ETF demand does not affect a stock's price on average, buy orders due to ETF rebalancing decisions have a strong positive effect on a stock's returns, and sell orders due to ETF rebalancing decisions have a strong negative effect on a stock's returns."*

Monthly Panel Regressions, All Stocks

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

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