ON INDEX INVESTING

Jeffrey L. Coles[a], Davidson Heath[a], Matthew C. Ringgenberg[a*]

[a] *David Eccles School of Business, University of Utah, 1655 Campus Center Drive, Salt Lake City, UT 84112 USA*

First Draft: March 2017
This Draft: January 2022[†]

ABSTRACT

We empirically examine the effects of index investing using predictions derived from a Grossman-Stiglitz framework. An exogenous increase in index investing leads to lower information production as measured by Google searches, EDGAR views, and analyst reports, yet price informativeness remains unchanged. These findings are consistent with an equilibrium in which investors choose to gather private information whenever it is profitable. As index investing increases, there are fewer privately-informed active investors (so overall information production drops), but the remaining mix of investors adjusts until the returns to active investing are unchanged. As a result, passive investing does not undermine price efficiency.

Keywords: index investing, information production, market efficiency, passive investing

JEL Classification Numbers: G12, G14

*Corresponding author: E-mail address: matthew.ringgenberg@eccles.utah.edu; Tel.: 801-213-6916

†Comments welcome. The authors thank an anonymous referee, Wei Wei and Alex Young for discussions about Russell research designs, Tony Cookson, Francesco Franzoni, Valentin Haddad, Olga Kolokolova, Will Mullins, Stefan Nagel, Toni Whited, participants at the 2017 Olin Business School Wealth and Asset Management Research Conference, the 2018 Financial Intermediation Research Society conference, the 2018 Santiago Finance Workshop, 2018 Tsinghua Finance Workshop, the 2018 Western Finance Association Annual Meeting, and seminar participants at Colorado State University, George Washington University, the University of Melbourne, University of Nebraska, and the University of South Florida. We thank Brendan Peek for excellent research assistance. We also thank Russell for providing index data. All errors are our own.

On Index Investing

First Draft: March 2017
This Draft: January 2022

ABSTRACT

We empirically examine the effects of index investing using predictions derived from a Grossman-Stiglitz framework. An exogenous increase in index investing leads to lower information production as measured by Google searches, EDGAR views, and analyst reports, yet price informativeness remains unchanged. These findings are consistent with an equilibrium in which investors choose to gather private information whenever it is profitable. As index investing increases, there are fewer privately-informed active investors (so overall information production drops), but the remaining mix of investors adjusts until the returns to active investing are unchanged. As a result, passive investing does not undermine price efficiency.

Keywords: index investing, information production, market efficiency, passive investing

JEL Classification Numbers: G12, G14

I. Introduction

Over the last two decades, the amount of capital devoted to index investing has grown by trillions of dollars (Bogle (2016)). This increase in passive index investing is not without controversy. Passive investors are necessarily free-riding on the research and analysis performed by active managers. This suggests a trade-off: while passive managers allow investors to earn index returns for low fees, the effort of active managers helps ensure that prices correctly reflect fundamental value. Put differently, not everyone can index, someone has to be *active*. The question is, does the rise of index investing change information production in the economy? If so, does it affect the informational efficiency of stock prices?

In this paper we empirically examine the effects of index investing on information production and price efficiency. We derive predictions using a simple extension to the Grossman and Stiglitz (1980) model of endogenous information acquisition and we test these predictions using Russell Index reconstitutions as a source of exogenous variation in passive investing.[1] Over our sample period from 2007 to 2016, we find that shocks to the mix of passive and active investors cause significant changes in trading behavior and information production. Nonetheless, consistent with the theoretical predictions from a class of equilibrium models, represented by Grossman and Stiglitz (1980) and our extension of that model, these shocks *do not* alter the information content of prices. Our results suggest that index investing does have effects on investment and information production. Nevertheless, the fraction and effort of informed investors adjust so that the relation between price and fundamental value is unchanged. In other words, the rise of index investing does not significantly affect market efficiency.

The theoretical literature provides no consensus on the relation between investor composition and market efficiency. Some theoretical models predict that a change in index investing does alter price efficiency. For example, Baruch and Zhang (2019), Bond and Garcia (2019),

[1]Wei and Young (2021) show that many existing studies that use Russell Index reconstitutions suffer from selection bias. Importantly, we develop a new identification strategy that avoids these issues. We discuss this in greater detail in Section III.

and Breugem and Buss (2019) all indicate that a first-order effect of an increase in index investors in the market should be a reduction in price informativeness.

In contrast, another class of models argues that in equilibrium the composition of investors does not necessarily affect price efficiency (e.g., Grossman and Stiglitz (1980), Liu and Wang (2019), and Davila and Parlatore (2019)). In the classic Grossman and Stiglitz (1980) model, investors choose whether to pay the cost of becoming informed. While their model predates the rise of index investing, it provides a framework for understanding the effects of a change in investor composition. We augment their model to include a choice by investors – investors choose to be passive, active and publicly-informed, or active and privately-informed.[2] The resulting framework predicts the rise of index investing will have *no* effect on price efficiency. The intuition is simple. In equilibrium, a decrease in the cost of index investing leads to more index investors and less active investors. However, because investors always choose to gather information when it is profitable, the mix of publicly-informed and privately-informed investors adjusts such that the returns to active investing remain unchanged even as index investing increases. As a result, price efficiency remains unchanged.

While the large and growing theoretical literature generates a diverse set of predictions, we provide the first empirical evidence that specifically tests these different predictions. In contrast to the numerous models suggesting index investing will alter price efficiency, our findings support the alternative class of models that predict there is no effect. Of course, null effects are difficult to establish. We conduct a variety of tests, first to verify that our experimental methods are adequately powered and, second, to isolate the economic mechanisms at work.

We start by examining the direct effects of index membership. After a switch in Russell Index assignment, we find strong evidence of a shift in the composition of investors. When a stock switches in to the Russell 2000 index from the Russell 1000, we find that ownership by

[2]The choice between publicly-informed and privately-informed represents the idea that investors choose how much to invest in costly information production and acquisition.

passive index funds increases by approximately 2% of its market capitalization. Ownership by active funds falls by a similar magnitude. The results confirm that our setting captures a meaningful shock to investor composition.

We then examine whether index investing affects investor behavior and stock price dynamics. We find that it does: more index investing causes stocks to have (i) higher share turnover, (ii) higher short interest, and (iii) higher correlation with index price movements. We also find that more index investing is associated with higher volatility, an out-of-sample replication of the findings of Ben-David, Franzoni, and Moussawi (2018). Moreover, our identification strategy allows us to separately estimate the effects of switching in *and* out of the Russell 2000 index; as expected, we find symmetric and opposite effects from switching *in* relative to *out* of the index.

We then investigate our two primary questions. First, we examine the effects of index investing on information production about the underlying stocks. In our extension to the Grossman and Stiglitz (1980) model, a decrease in the total mass of investors that pay to gather information leads to less information production in aggregate. Using three different measures of information production – Google search volume, EDGAR page views from the Securities and Exchange Commission (SEC), and buy-side analyst reports – we find that more index investing leads to less information production about individual index stocks. Following an exogenous increase in index fund ownership, Google search volume falls by 3.8%, EDGAR page views fall by 14.1%, and the number of analyst reports falls by 10.8%.

Second, we examine whether the change in information production affects price efficiency at the stock level. While the prior literature has also documented evidence of various effects on asset markets (like changes in correlations and volatility), to date there is less evidence on price informativeness. We find that index investing does not affect price informativeness. Our estimates, which are well-powered, meaning that there is a high probability that our test method will find an effect if there is an effect to be found, fail to reject the null of no change for a range of measures of price informativeness – specifically variance ratios, anomaly mispricing, and post-earnings announcement drift. Thus, after an exogenous increase in index investing, information production about individual assets falls, but there is zero change in the informational efficiency of prices.

Taken together, our findings are consistent with the idea that index investing changes the composition of the investor base, which alters market dynamics and information production, but that active investors endogenously respond to ensure that price efficiency is unchanged. Several models predict that index investing will have first-order effects on price efficiency (e.g., Baruch and Zhang (2019), Bond and Garcia (2019), and Breugem and Buss (2019)). Our precisely estimated non-results are inconsistent with that hypothesis. Instead, our evidence accords with our extension of Grossman and Stiglitz (1980) and related work such as Liu and Wang (2019) and Davila and Parlatore (2019). In these models, changes in the number of active versus index investors or the float of traded shares have no first-order effect on price efficiency because the fractions of publicly-informed and privately-informed active traders adjust endogenously so that the payoff to gathering information does not change.[3] To the best of our knowledge we are the first to directly test this prediction from Grossman and Stiglitz (1980).

Importantly, our non-results on price efficiency are not due to low statistical power. First, we note that our research design does detect multiple direct effects on fund holdings and underlying asset markets. Second, to examine statistical power directly, we calculate the minimum detectable effect size ("MDES") for each of our estimates (Bloom, 1995).[4] The MDES gives a measure of the magnitude of effects that our tests could reliably detect, if they were present. The MDES for our tests varies between one-third and one-sixth of one sample standard deviation for our outcome variables of choice, demonstrating that our tests are not underpowered. Put differently, our tests do have the power to detect relatively small

[3]Importantly, this result holds for *any* size change in the quantity of index investors. Our results, which examine a relatively large change in the quantity of index investors equal to 1-2% of shares outstanding each year, are consistent with this prediction.

[4]We assume in each case that the main treatment coefficient has a t-distribution with the appropriate degrees of freedom, and compute the minimum detectable effect size under a two-sided test with $\alpha = 0.05$ and $\beta = 0.2$. See Section IV for details.

effects: index investing simply does not change price efficiency.

Of course, one of the key challenges to understanding the effect of index investing is that the quantity of index capital allocated to a stock is not random. To address this, we develop and apply a new research design based on post-2007 Russell index reconstitutions.[5] Starting in 2007, Russell Investments changed the methodology used to create its well-known Russell 1000 and 2000 indices. These changes eliminated the sharp discontinuity between stocks in the Russell 1000 and 2000, rendering traditional regression discontinuity designs infeasible.[6] We develop a new methodology that exploits post-2007 Russell index reconstitution in order to generate exogenous variation in index investing.[7,8]

Overall, our paper makes several contributions to the literature on information economics. A number of academics and practitioners have questioned whether society spends the correct amount on information production and acquisition,[9] and the rise of index investing has generated more questions on this important issue. We provide novel evidence that market participants adjust to an increase in index investing, such that the economic returns to information acquisition are unchanged. We also confirm important predictions from the Grossman and Stiglitz (1980) model.

We document a number of novel empirical facts. We show clear evidence that index re-balancing causes index investors to replace active managers as owners in a stock. We then measure and assess the implications of these changes for asset prices. Consistent with several recent papers, we do find evidence that index investing changes market dynamics. Our results show that higher index investing is associated with higher turnover, higher short interest,

[5]The research design we develop here is also used in Heath, Macciocchi, Michaely, and Ringgenberg (2021).

[6]A number of papers exploit the discontinuity from Russell reconstitutions using data *before* this 2007 change was implemented (e.g., Mullins (2014), Chang, Hong, and Liskovich (2015), Appel, Gormley, and Keim (2016), Crane, Michenaud, and Weston (2016), Schmidt and Fahlenbrach (2017)).

[7]Underway in the literature is an active discussion of the appropriate empirical approach. See Ben-David, Franzoni, and Moussawi (2019) and Appel, Gormley, and Keim (2020), for example.

[8]Heath, Ringgenberg, Samadi, and Werner (2019) show that natural experiments (such as Russell reconstitutions) are often re-used leading to multiple testing problems. Because our research design is specific to the post-banding period, our tests are in effect out-of-sample to the existing Russell literature. Moreover, since our primary findings are non-results, multiple testing adjustments do not change our inferences.

[9]See, for example, Budish, Crampton, and Shim (2015) and Fraser-Jenkins et al. (2016).

and higher index correlations. Nonetheless, using the same empirical strategy, we find no evidence that index investing changes price informativeness. Shocks to index investing do not change variance ratios, post-earnings announcement drift, or anomaly mispricing.

We explore the mechanism behind this result. Consistent with endogenous information acquisition, we document novel evidence that index investing leads to a drop in aggregate information production. As index investing increases, we find decreases in Google search volume, EDGAR page views, and buy-side analyst reports, and vice versa. Nevertheless, we find no evidence that index investing affects price efficiency. Taking all the evidence together, the main conclusion becomes clear: Index investing shifts the composition of investors, which leads to changes in trading behavior and information production, but does not affect the ability of arbitrageurs to impound information into prices. As such, our findings present novel evidence confirming key predictions in the original and augmented versions of Grossman and Stiglitz (1980).[10]

Finally, our paper also contributes a new methodology to examine the impact of index reconstitutions. While a number of papers have used index reconstitutions as part of an identification strategy (e.g., Boone and White (2015), Crane et al. (2016), Appel et al. (2016), etc.), several papers show that the approaches in these papers may lead to selection bias (Wei and Young (2021), Glossner (2021)). In section III.C, we show evidence that our methodology does not have issues with selection bias or low statistical power.

II. Background

In this section, we briefly discuss existing work on index investing and its effect on market outcomes. We then provide a theoretical framework that builds on the model in Grossman and Stiglitz (1980) to examine the possible effects of index investing.

[10]Sunder (1992) presents evidence from laboratory experiments that supports predictions from the Grossman and Stiglitz framework. To the best of our knowledge, our paper contains the first empirical evidence on this point.

A. Related Literature

While index investing has existed for nearly 40 years (Bogle, 2016), much of the literature on the relation between index investing, market efficiency, and other market characteristics is relatively recent. The results vary. Comparative statics analysis of an increase in the proportion of investors who can be understood to be passive can lead to non-monotone (Goldstein & Yang, 2017) or ambiguous - depending on the model parameters - (Stein, 1987; Subrahmanyam, 1991) effects on market efficiency. Baruch and Zhang (2019), in a rational expectations framework, offer the more definitive statement that an increase in the proportion of investors who index causes a decrease in market efficiency, insofar as the statistical fit of the CAPM regression decreases and the conditional variance of payoff increases. Cong and Xu (2016) show that the presence of and trading in composite securities, such as index ETFs, can lead to lower asset-specific information in security prices.

Other models partially or fully endogenize the extent to which investors choose to be active versus passive. Brown and Davies (2017) show that downward pressure on the fees (e.g., due to increased index investing) of a fixed set of active managers can reduce fund manager effort and the informational content of prices. Breugem and Buss (2019) show that an increase in the share of investors who benchmark to an index can limit investors' willingness to speculate, which reduces the value of private information. In equilibrium, investors acquire less information and informational efficiency declines. Likewise, in the Bond and Garcia (2019) model as indexing becomes cheaper indexing increases, individual stock trading decreases, and aggregate price efficiency falls. Haddad, Huebner, and Loualiche (2021) examine an equilibrium model of investor competition. Following a shift in the composition of investors, their model shows that equilibrium outcomes may or may not change, depending on how competitive the market is.[11]

Empirically, several recent papers have investigated the market impact of investing in ETFs. Perhaps most closely related to our study, Ben-David et al. (2018) examine firm volatility using Russell Index reconstitutions to generate variation in index ownership. They find that an increase in index investing is associated with higher volatility and more negative correlation in stock prices over 1996 to 2006. Israeli, Lee, and Sridharan (2016) find that higher ETF ownership is associated with higher future earnings response coefficients and decreased liquidity for stocks with higher ETF ownership. In contrast, Glosten, Nallareddy, and Zhou (2016) find that an increase in ETF ownership is associated with more accounting information in market prices. Our experimental setting is different: because we examine changes in ownership composition that result purely from Russell reconstitutions, our results are driven by variation in the proportion of passive investment only. As such, our paper investigates different questions.

Second, Israeli et al. (2016) and Glosten et al. (2016) focus primarily on the incorporation of accounting information into stock prices. In contrast, we apply variance ratio tests to assess weak form efficiency. We also examine anomaly mispricing. To the best of our knowledge, our paper is the first to use exogenous variation in passive ownership to identify the impact of index investing on the information content of prices and on the trading behavior of institutional traders. In Section IV, below, we present results from a Durban-Watson-Hausman Chi-squared test which shows strong evidence that index investing and price efficiency are endogenously determined. Accordingly, we believe our setting is an ideal laboratory for examining the impact of index investing.

Finally, our empirical design makes use of the mechanical rules for Russell index assign-

[11] The theoretical literature addresses other implications of the mix of passive versus active investors. Examples of effects in markets that pertain to price informativeness include: trading and liquidity (Subrahmanyam, 1991; Basak & Pavlova, 2013; Cong & Xu, 2016); asset correlations (Basak & Pavlova, 2013; Cong & Xu, 2016; Baltussen, van Bekkum, & Da, 2019; Baruch & Zhang, 2019); price and return volatility (Stein, 1987; Breugem & Buss, 2019; Basak & Pavlova, 2013; Chabakauri & Rytchkov, 2021); the Sharpe ratio (Baruch & Zhang, 2019); market reversals (Bond & Garcia, 2019); and risk-sharing and welfare (Stein, 1987; Bond & Garcia, 2019).

ment, as have Ben-David et al. (2018) and a few other recent papers.[12] Because Russell changed their methodology for index reconstitutions starting in 2007, our sample period requires us to develop a different methodology to identify changes in index membership. Section III describes our method.

B. Theoretical Predictions

Previous studies have documented that index investing affects the market dynamics of underlying stocks (i.e. Ben-David et al. (2018)). In this paper, our goal is to examine the effects of index investing on the information content of prices. To motivate our empirical tests and put our estimates in context, we develop a simple model that builds on the classic Grossman and Stiglitz (1980) model. While the Grossman and Stiglitz (1980) model does not examine index investing, it provides a useful framework for understanding how information is incorporated into asset prices in equilibrium. We expand the Grossman and Stiglitz (1980) model to include index investors. Specifically, the model includes three types of investors: passive index investors, active publicly-informed investors, and active privately-informed investors.

Consider a single risky asset with payoff $\tilde{\theta} + \tilde{\epsilon}$ one period from today, where $\tilde{\theta}$ is the fundamental value. At $t = 0$, a total of $\tilde{x}$ shares of the asset are offered by price-insensitive noise traders. $\tilde{\theta}$, $\tilde{\epsilon}$, $\tilde{x}$ are jointly normal and uncorrelated with variances σ_θ^2, σ_ϵ^2, and σ_x^2.

There is a unit mass of atomistic investors who have CARA utility with risk aversion ψ. Each investor chooses between three options: To be a passive index investor, an active publicly-informed investor, or an active privately-informed investor. The fixed cost of being an index investor is c_I; the fixed cost of being an active investor is c_A. The fraction of investors

[12]For example, to address the extent to which demand curves for stock slope downward, Chang et al. (2015) use Russell reconstitutions to measure the price effects from index additions and deletions. Mullins (2014) uses Russell reconstitutions to examine the impact of institutional ownership on corporate governance. Appel et al. (2016) find that higher passive investment was associated with better governance and changes in the type of campaigns launched by activist investors. Crane et al. (2016) find that higher passive investment was associated with higher payout to investors. Finally, Schmidt and Fahlenbrach (2017) find that higher passive investment was followed by increases in CEO power and worse M&A outcomes.

who choose to index is denoted by m. Index investors choose their optimal allocation (X) to the asset ex ante:

$$X_{Index} = \frac{1}{\psi} \frac{E[\tilde{\theta} - \tilde{P}]}{\sigma^2_{Index}} \tag{1}$$

where $\tilde{P}$ is the asset's price at $t = 0$ and σ^2_{Ind} is the unconditional variance of the asset's payoff. As a whole, indexers demand mX_{Ind} shares of the asset regardless of its price – that is, indexers are price-insensitive. Thus the net supply of shares that is offered to active investors is $\tilde{x} - mX_{Ind}$.

The remaining mass of investors, 1-m, are price-sensitive active investors. Each active investor chooses whether to pay an additional information cost c to learn the nonpublic signal θ. If they do, they are a privately-informed active investor and have demand for the asset:

$$X_I = \frac{1}{\psi} \frac{\theta - P}{\sigma^2_{\epsilon}} \tag{2}$$

If they choose not to pay c, they are a publicly-informed active investor. They have demand for the asset:

$$X_U = \frac{1}{\psi} \left(\frac{E[\tilde{\theta}|P] - P}{Var[\tilde{\theta}|P] + \sigma^2_{\epsilon}} \right) \tag{3}$$

The fraction of active investors that choose to pay c to become privately-informed is denoted λ. Thus, the mass of privately-informed active investors is $(1-m)\lambda$, and their total expenditure on information production is $(1-m)\lambda c$.

In equilibrium, each investor is indifferent between the three categories:

$$E\left[-exp(-\psi \tilde{W}_1)|X_{Index}\right] = E\left[-exp(-\psi \tilde{W}_1)|\tilde{X}_I\right] = E\left[-exp(-\psi \tilde{W}_1)|\tilde{X}_U\right] \tag{4}$$

That is, the ex-ante expected CARA utility of final wealth for an index investor, a publicly-informed active investor, and a privately-informed active investor is the same. We solve for the equilibrium numerically (details are in Appendix 2).

As in Grossman and Stiglitz (1980), we find that the correlation ρ_θ between price and fundamental value equals:

$$\rho_\theta = \sqrt{1 - \frac{\sigma_\epsilon^2}{\sigma_\theta^2}(e^{2\psi c} - 1)}. \tag{5}$$

Note that the cost advantage of indexing, $c_I - c_A$, and the mass of investors who choose to index, m, are both absent in this formula. Thus, both the cost advantage of indexing and the extent to which investors choose to index are irrelevant to the information content of the market price.

To sum up: Privately-informed active investors pay $c_A + c$ and condition their trading on both θ and P. Publicly-informed active investors pay c_A and condition their trading on P alone. Index investors pay c_I and do not condition their trading on any asset-specific information. The information content of the price is determined by the choice by active traders whether to pay to observe the private information, which in turn is determined by three parameters: (i) the information acquired relative to the uncertainty of the asset's payoff, (ii) the cost of information acquisition, and (iii) their risk aversion. Importantly, no other model parameters appear in Equation (5). In particular, the mass of index investors in the asset, m, is *irrelevant* to the information content of the price. Using these equilibrium conditions, we next use the model to assess how a change in the cost advantage of index investing affects price efficiency, information production, and the fraction of investors who choose to become active versus passive.

B.1. Comparative Statics

It is well established that index funds have substantially lower fees on average than active funds (Malkiel, 2013). In response to low-fee index funds, active funds have also lowered their fees (Cremers, Ferreira, Matos, and Starks (2016)). While this competitive response could dampen the cost advantage of index funds, the costs of indexing have continued to fall over time and in many cases, fees for index investing have become so low they are nearly zero - for example, the expense ratio for the Vanguard S&P500 ETF is now 0.03 percent per year. We model the "rise of index investing" as a consequence of a decrease in the cost of index investing c_I relative to the cost of active investing c_A. Figure 1 shows how the market equilibrium changes as the cost advantage of indexing increases.

Figure 1 Panel A plots the fraction of investors who choose to index. When the cost advantage of indexing is sufficiently low, no investors choose to index. This is because publicly-informed active investors outperform indexers in expectation, since they condition their investment on the price relative to fundamental value. This prediction is also consistent with the argument in Pedersen (2018) that active investors could outperform index investors because index investors bear the cost of price pressure when indexes are rebalanced or new shares are issued. However, when the cost advantage of indexing passes a threshold, a positive and increasing fraction of investors choose to index.[13]

We next examine how the rise of index investing affects information production and asset prices.

Remark 1 *The volatility of the asset's price is increasing in the fraction of index investors.*

Remark 1 follows directly from the model derivation in Appendix 2.A.4, and motivates our first empirical tests on the effects of index investing. Following an exogenous increase in the cost advantage to index investing and an associated increase in the proportion of index investors, we expect to see an increase in asset-specific volatility.

Remark 2 *An exogenous increase in index investing does not change the fraction of active investors who choose to gather information.*

[13]As we reduce c_I, the cost advantage of indexing eventually becomes so high that 100% of investors choose to index and the model breaks down. Thus, we analyze the effects of reducing c_I starting from c_A up to the point at which 99% of investors choose to index (the open circle at the end of each line in Figure 1). In reality, it seems unlikely that the market would consist only of index investors. While our model does not explicitly consider equilibrium in the market for active investing services, our argument here suggests that Figure 1 contains the applicable domain of indexing cost advantage and the corresponding range of equilibrium outcomes in the asset market.

What happens to the choice to gather information? Figure 1 Panel B plots the fraction of active investors who choose to become privately-informed, λ. Even as index investing becomes more prevalent, the fraction of active investors who choose to become privately-informed is constant across the entire range of equilibrium outcomes. This is one of the key results: as the cost advantage of indexing rises, the mass of indexers rises and the mass of active investors falls. The *fraction* of active investors choosing to gather information, however, stays constant.

Next, we examine how this affects information production in the economy and price informativeness. Panel C plots the total production of asset-specific information. Because the mass of active investors falls as index investing increases, there are fewer investors producing information. Thus, the total production of asset-specific information falls.

Remark 3 *An exogenous increase in index investing leads to a drop in asset-specific information production.*

Remark 3 motivates our empirical analyses on information production: Following an increase in index investing, we expect to see fewer investors gathering firm-specific information. The intuition is simple: there are fewer active investors who choose to become privately-informed, so there are fewer investors gathering information. For example, we expect to see a decrease in the number of investors downloading firm-specific financial information from the SEC EDGAR database and a decrease in the number of analyst reports produced.

Finally, we examine price informativeness. Panel D of Figure 1 shows the key and perhaps surprising result of the model: while the total quantity of information production falls, fundamental price informativeness, ρ_θ, *does not change* for any level of index investing.[14] In equilibrium the informativeness of the price is determined by the *ratio* of active privately-informed investors, who correct mispricing, to active publicly-informed investors who generate mispricing. While an increase in the quantity of index investors leads to less active investing overall, as shown in Panel B, the ratio of active privately-informed to active publicly-informed investors is constant for any level of index investing. As a result, the quantity of index investors does not matter because the remaining investors will adjust their behavior until price informativeness is unchanged.

Remark 4 *An exogenous increase in index investing leads to no change in price informativeness.*

Remark 4 motivates our empirical analyses on price informativeness: following an increase in index investors, we expect to see no change in price informativeness. Again, the intuition is simple: investors will always choose to become privately-informed active traders if it is profitable to do so. Importantly, this result holds for any size change in the quantity of index investors (i.e., for small or large increases in index investing we expect to see the same thing – no change in price informativeness).

Of course, our model is necessarily a simplified depiction of the market. In reality, funds might not perfectly fit into one of our three investor categories: passive, publicly-informed active, and privately-informed active. For example, the Sovereign Wealth Fund of Norway is one of the largest investors in the world, and it invests in global equities using a strategy that is based on a benchmark index, but allows for discretion to deviate from the index.[15] Similarly, Robertson (2019, 2020) shows that some index funds are benchmarked to an index that they created. As such, it is not clear that such index funds are truly passive, since they

[14] This result holds for the domain of cost parameters under which we are able to discern an equilibrium. If the cost advantage to index investing becomes extravagantly high, at which point all investors index, neither have we determined whether there exists an equilibrium, nor, conditional on existence, have we identified the nature of equilibrium. Of course, such a high cost differential accords with the intuition that the stock market potentially would depart significantly from semi-strong-form efficiency. On the other hand, such an outcome would require an unlikely equilibrium in the market for active investing services, in which the marginal cost to the very first investor of becoming active exceeds the benefit of doing so.

[15] The fund states, "We keep the fund close to the benchmark, but all of our investment strategies also have active components. Norges Bank Investment Management (2022).

have discretion to change the index members and weights. Thus, index investing may not be a completely separate alternative to active investing in practice, investors might do both. We also note that index investing may have other effects that are not discussed in our model. For example, Schmidt and Fahlenbrach (2017) show evidence that index investing can affect corporate governance, and as a result, firm value. While these issues are important, they are outside the scope of our model and we believe the general point of the model stands: investors will choose to invest more in information production whenever it is profitable to do so. As a consequence, the rise of index investing should not significantly alter price efficiency.

III. Data and Research Design

To empirically test the predictions developed above, we examine the effects of index investing on information production and price efficiency using variation in Russell Index membership as an exogenous shock to ownership by index funds. To do this, we combine stock data from the Center for Research in Security Prices (CRSP) with firm data from Compustat, ownership data from Thomson Reuters S12 and 13F, and Russell index data directly from Russell Investments.

A. Data

Russell index membership data come directly from Russell. Financial data on our sample of stocks are from CRSP and the merged CRSP-Compustat database.

From Compustat, we also obtain information on firm-level short interest. Short interest data are compiled by the stock exchanges twice per month, near the middle and end of each month, and the data are publicly released four business days later. Prior to September of 2007, short interest data were only reported once per month, near the middle of the month, so for consistency we use mid-month short interest in all of our analyses.[16] We also use information on anomaly mispricing using the mispricing score from Stambaugh, Yu, and Yuan (2015), available on Robert Stambaugh's website. We obtain daily search volume for each stock ticker from Google Trends. We obtain data on analyst coverage from IBES, and data on EDGAR search volume from the SEC EDGAR log files.

We classify funds as active or passive using the CRSP index fund flag (index fund flag = "D"). This definition includes open-ended mutual funds, closed-end funds and ETFs (which we refer to collectively as "funds"); a fund is designated as an index fund if its objective is to replicate an index and not to generate abnormal returns. Neither our model nor our tests make any assumptions about the *form* of passive investing (i.e., index funds can be mutual funds, closed-end funds, or ETFs).

Fund holdings data come from the Thomson Reuters S12 and CRSP databases. We compute the ownership of each sample stock by every fund in each quarter, using the union of the two datasets: When one of the two holdings datasets records fund X as holding N shares of firm Y in quarter t, we include it unless the other holdings dataset records fund X holding M > N shares of firm Y in quarter t. We use the number of sole-voting shares held where available, otherwise the total shares held.

We also compute ownership by institutional investors more broadly using the Thomson Reuters 13F database, which captures large institutional investors including investment advisors, banks, broker-dealers, insurance companies, hedge funds, mutual funds and pension funds.

Our measures of ownership for each stock i as of the end of each quarter t are defined below. All are expressed as a percent of the stock's total market capitalization.

- $FundOwn_{it}$: Total ownership by all funds in the union of the S12 and CRSP fund

[16] We compute the short interest ratio in two ways for each firm and month: (i) we divide short interest by shares outstanding from CRSP and (ii) we divide adjusted short interest by adjusted shares outstanding from CRSP, where the adjustment accounts for splits, buybacks, etc. These two measures should be identical, so we drop observations when the absolute deviation between the two measures exceeds 10%.

holdings databases.

- $FundOwn_t^{Passive}$: Total ownership by all passive funds as defined above.
- $FundOwn_{it}^{Active}$: Total ownership by all active funds as defined above.
- $FundOwn_t^{P,R2000}$: Total ownership by all passive funds as defined above and whose target index was the Russell 2000.
- $FundOwn_t^{P,R1000}$: Total ownership by all passive funds as defined above and whose target index was the Russell 1000.
- $InstOwn_t$: Total ownership by all institutional investors in the 13F database.

Our sample consists of all stocks in +/-100 rank windows around the upper and lower bands, each year from 2007 to 2016, that were *potential switchers* (i.e. stocks near the upper band that were in the Russell 2000 in May of the cohort year, and stocks near the lower band that were in the Russell 1000 in May of the cohort year). Figure 3 shows the sample stocks for the first post-banding reconstitution in June 2007. To ensure that poor liquidity or market micro-structure issues do not affect our estimates we drop stocks that had a May share price under $5 per share (Asparouhova, Bessembinder, & Kalcheva, 2013). Our results are similar if we omit this filter.

Table I presents summary statistics, measured as of May prior to index reconstitution, for both the upper-band and lower-band samples. The firms appear similar in nearly every dimension, including their passive and active ownership levels, turnover, volatility, short interest, and a variety of measures of information production (Google search volume, EDGAR downloads, and analyst reports). The one notable difference is market capitalization, which is the variable on which Russell creates the two indexes. On average, upper-band firms are twice as large as the lower-band firms (although the magnitudes of $3B versus $1.6B market capitalization are not that large relative to the standard deviations of $1 billion and one-half billion, respectively). Aside from that difference, the two samples appear very similar across all variables. That is, they are two narrow slices of mid-cap U.S. stocks that are similar on other observable characteristics.

B. Russell Index Assignment

In June of each year Russell Investments reconstitutes their popular Russell 1000 and 2000 indexes. To determine index membership, Russell ranks qualifying U.S. common stocks by their market capitalization as of the last business day in May. Prior to 2007 index assignment followed a simple threshold rule: stocks ranked from 1-1000 were assigned to the Russell 1000 while stocks ranked from 1001-3000 were assigned to the Russell 2000.

Starting in June 2007, Russell implemented a new assignment regime ("banding"). After sorting stocks by their market capitalization, Russell computes an upper and lower band around the rank-1000 cutoff; the bands are calculated as +/- 2.5% of the total market capitalization of the Russell 3000E.[17] Stocks within the bands do not switch indexes. That is, if a stock that is currently in the Russell 2000 is above the rank-1000 cutoff but below the upper band, it will stay in the Russell 2000 the following year, and vice versa.

Figure 2 plots index assignments in June 2007, the first year of the banding regime. The change eliminated the discontinuity at the rank-1000 threshold, however the banding regime replaced it with two *new* discontinuities at the upper and lower band conditional on a stock's previous index assignment. For example, consider a stock in the Russell 2000 that is nearby the upper band when the indexes are reconstituted. The stock's index assignment depends on four parameters as calculated by Russell: 1) The stock's ranking in the Russell 3000; 2) The market capitalization of the rank-1000 stock; 3) The total market capitalization of the Russell 3000E; 4) The cumulative market cap of the stocks ranked below the focal stock but above the rank-1000 stock. All four parameters are difficult to predict ex ante – indeed, Russell does not make their unadjusted market capitalizations or rankings available *ex post*.

[17]The 3000E is an "extended" version of the Russell 3000 that includes microcap stocks.

All four parameters are difficult or impossible to manipulate. This line of reasoning suggests that within a sufficiently narrow window around each band in each year, whether a stock ranks above or below the band – and therefore switches or stays – is as good as randomly assigned.

C. Research Design

For each Russell reconstitution since 2007, we select a *cohort* containing two sets of treated and control stocks. We select all stocks that were potential switchers, based on their lagged index membership, in windows of +/-100 ranks around the upper and lower band. Consider for example two stocks A and B that are similar in every aspect, including that both are in the Russell 1000 index in the year prior to treatment. Both stocks experience negative returns in the year prior to treatment and drop in the rankings. Firm A's market capitalization falls by 12% while Firm B's market capitalization falls by 12.1% and as a result, stock A stays in the Russell 1000 (and is a control), whereas stock B crosses the lower band and switches to the Russell 2000 (and is treated). Thus our identification strategy compares stocks that started in the same index and are similar in every dimension – including their lagged returns – except that they landed by a small margin on different sides of the same band.

Figure 3 shows the treated and control stocks around both bands in the 2007 cohort. Not all stocks within the +/-100 rank window around each band were potential switchers. In both cases, about half of those stocks ultimately switched indexes, which again suggests that within our selected cohorts, index assignment was as good as randomly assigned. Detailed balance tests (Appendix Table A1) show similar evidence, and our results are robust to varying the window size. Figure 3 also underlines that our sample is not selected on ex-post results (i.e., it is not based on whether a firm switches indexes each year). Rather, to ensure balance ex-ante, sample firms are selected within a fixed +/100 rank window around the upper and lower bands prior to index reconstitution. Because stock returns are positive on average, the number of stocks that land in the window around the upper band tends to be higher than the number of stocks that land in the window around the lower band. Thus, on average, there are more stocks in the upper-band sample than the lower-band sample each year.

For each stock in each cohort, we include quarterly or monthly observations, depending on the measure, for one year before and one year after assignment. We estimate difference-in-differences specifications, with fixed effects for each stock in each cohort. For the upper-band treatment effect, the specification is:

$$\begin{aligned} Y_{jct} = \beta\, I\{R2000 \rightarrow R1000_{jc}\} \times PostAssignment_{ct} \\ + \phi_{jc} + \lambda_t + \epsilon_{jct}, \end{aligned} \tag{6}$$

and for the lower-band treatment effect the specification is:

$$\begin{aligned} Y_{jct} = \beta\, I\{R1000 \rightarrow R2000_{jc}\} \times PostAssignment_{ct} \\ + \phi_{jc} + \lambda_t + \epsilon_{jct}, \end{aligned} \tag{7}$$

where ϕ_{jc} and λ_t are stock-by-cohort and time fixed effects. The firm-by-cohort fixed effects sweep out any time-invariant differences between treated and control stocks. Importantly, this includes preexisting differences on any characteristics between treated and control stocks. For example, by definition stocks that switched into the Russell 1000 had a higher lagged return and larger market cap on average than stocks that did not. Any such preexisting differences cannot explain our findings, or be explicitly controlled for in the regression, because they are already swept out by the stock-by-cohort fixed effects. Likewise, the time fixed effects remove any aggregate trends in stock market activity or average ownership by any investor category. While many other papers that examine Russell Index reconstitutions use an instrumental variables (IV) approach, we use a difference-in-differences approach.

This has several advantages. IV approaches crucially require an untestable only-through condition, that all causal effects they identify must result directly from the instrumented relation estimated in the first stage. Our approach is more flexible, as it compares treated and control firms before and after index reconstitution. Our estimates in Table II provide evidence that Russell reconstitutions do change ownership, but our estimates of the effects on price efficiency and information production do not depend on the measurement of these ownership changes. Put differently, our estimates on the effects of index investing reflect the impact of every investor who changes their holdings as a result of index reconstitution. This research design also has advantages over a regression discontinuity design (RDD), including that our estimates are not sensitive to measurement error in the forcing variable $CapRank$.[18] To check for selection bias, we run balance tests across a wide variety of firm variables measured prior to treatment. Appendix Table A1 presents the results, which uniformly find that our treated and control groups are indistinguishable *ex ante*. Thus, the balance tests suggest that our estimates are not contaminated by selection bias.

IV. Results

In this section we examine the effects of Russell index assignment on institutional ownership, trading behavior, and asset prices. We first examine whether, and how, index rebalancing affects the mix of active versus passive fund ownership. We then examine the direct effects of index investing on asset markets. Finally, we examine the relation between index investing and i) price informativeness and ii) trading by arbitrageurs. Overall, our findings suggest that higher index investing leads to significant changes in investor composition and information production, but index investing does not affect price efficiency.

[18] We thank Toni Whited for this point. Heath et al. (2021) discusses this issue in more detail.

A. Identification and Interpretation

In contrast to some existing papers, we do not use index assignment as an instrumental variable for passive fund investment, because that approach requires an only-through condition that any subsequent effects are only due to changes in passive fund investment as estimated in a first-stage regression. By contrast, our approach is a difference-in-differences analysis that compares treated and control firms before and after index reconstitution. This approach estimates the effects of index switching whatever they are.

The effects of index switching on passive fund ownership are symmetric and robust in our sample of funds (Table II). However, there are at least two other types of effects that will occur at the same time. First, there will be an effect on ownership by other non-fund passive investors who track the Russell indices. These effects bolster the plausibility of our overall conclusions, since they suggest a larger effect on passive investing in the same direction as we measure.

Second, there will be other effects not due to passive investing. For example: 1. Suppose there is a group of equity analysts that devote equal time to all R1000 stocks. A stock that switches into the R1000 will have lower passive investing and greater analyst coverage, but the latter effect is not driven by the former. 2. Suppose a small-cap active fund is benchmarked to the Russell 2000 and only looks at those stocks to identify opportunities. When a stock switches into the R2000, its expected holdings by that fund and expected active-fund attention and activity will go up, even as passive investing in that stock also goes up.

Thus one caveat to our empirical results is that the effects of index switching, whatever they are, is all that we can estimate without added assumptions that may or may not be justified or even testable.

B. Statistical Power and the Minimum Detectable Effect Size

In order to generate a clean comparison between treated and control stocks, our research design conditions down to a small subset of stocks that are close to the yearly Russell bands. One potential consequence is that our estimates could lack power. To examine this possibility, in our regression tables we report the *minimum detectable effect size* as defined by Bloom (1995).[19] We report the MDES in the original units of the outcome variable, and scaled by the sample standard deviation of the outcome variable. Doing this gives the reader a way to evaluate statistical power for each estimate and each outcome variable individually.

C. Effects on Fund Ownership

Table II presents difference-in-differences estimates for ownership by funds (open-ended mutual funds, closed-end funds, and ETFs), which compare the post-assignment changes in fund ownership for treated versus control stocks (stayers versus switchers) across the two yearly bands.

In both panels, we see that total fund ownership ($FundOwn$) in a stock does not change significantly across the Russell discontinuities. Assignment to the Russell 2000, however, strongly alters the *composition* of fund ownership. Ownership by passive index funds $FundOwn^{PASSIVE}$ increases by 2.09% of market capitalization, on average, for stocks that cross the lower band and switch into the Russell 2000. This increase is almost entirely driven by an increase in ownership by passive funds that track the Russell 2000. On the other hand ownership by passive funds falls by 0.96% of market capitalization, on average, for stocks that cross the upper band and switch into the Russell 1000. Interestingly, we also see evidence that passive funds are transacting with active funds on net (rather than retail traders or

[19] We assume in each case that the main treatment coefficient has a t-distribution with the appropriate degrees of freedom, and compute the minimum detectable effect size under a two-sided test with $\alpha = 0.05$ and $\beta = 0.2$. In practice, for N larger than 100, the MDES is 2.8 times the standard error of the coefficient. Intuitively, setting $\alpha = 0.05$, an effect that is 1.96 times the standard error would be detected 50% of the time; an effect that is 2.8 times the standard error would be detected 80% of the time corresponding to power $\beta = 0.2$. Note also that since our standard errors are HAC-robust and clustered by stock, the MDES inherits these properties.

uncategorized investors). In both panels, ownership by active funds $FundOwn^{Active}$ changes in the opposite direction, offsetting the rise in passive fund ownership.

The changes in holdings by passive funds are strongly significant because holdings by passive funds at the stock level are stable year to year. By contrast, although the point estimate of the change in holdings by active funds is of a similar size, the standard error is much larger because holdings by active funds are much more volatile.

Other institutional investors such as hedge funds, pension funds or sovereign wealth funds may be partially or heavily indexed and also react to a change in a stock's index assignment. To examine the broader effects of index switching on ownership, in column 6 we estimate the effects on ownership by all institutional investors that filed a 13F statement. In Panel A we see that switching from the Russell 1000 to the Russell 2000 is followed by an increase in institutional investor ownership of 5.4% of a firm's market cap on average. Because we have fewer observations near the lower band, this estimate has relatively low power. However, In Panel B we see that switching from the Russell 2000 to the Russell 1000 is followed by a decrease in institutional investor ownership of 3.8% of a firm's market cap, and this change is statistically significant at the 1% level. Thus, index assignment has effects on ownership by institutional investors more broadly. To the extent that multiple categories of indexed investors react to index switching, this suggests that our results generalize from the fund market and that both the causes and the effects of index investing apply more broadly.

How do the magnitudes of these changes compare to the overall level of passive investing? The average change in passive ownership for stocks that switch indexes is approximately 1.5% of the stock's total market capitalization. The average level of ownership by passive funds prior to treatment is approximately 10% of market cap (see Table I). Thus, the change represents roughly 15% of the baseline level of ownership by passive funds.

We conclude that the treatment effects of index switching are sufficiently large that our research design is plausibly adequately powered to conduct further tests of the effects on asset markets.

D. Effects on Underlying Asset Markets

Several recent theoretical models suggest that index investing could change the dynamics of stock markets and investor behavior, including changes to the correlation structure of prices (e.g., Baruch and Zhang (2019), Basak and Pavlova (2013)), and changes to the information gathering and behavior of informed investors (e.g., Cong and Xu (2016), Brown and Davies (2017), Bond and Garcia (2019)). In contrast, a long line of theoretical models on information production and informed trading behavior dating back to Grossman and Stiglitz (1980) suggests that such changes need not have an effect, in equilibrium, on the information content of asset prices. Using our new Russell research design, we explore the effects of index investing on a variety of stock-level outcomes suggested by theory during the period from 2007 to 2016.

Table III examines the direct effects of index investing on the markets for individual index stocks. Panel A shows the effects of switching into the Russell 2000, estimated within the lower-band sample only; Panel B shows the effects of switching out of the Russell 2000, estimated within the upper-band sample only. Panel A Column 1 shows that switching from the Russell 1000 to the Russell 2000, which causes an increase in passive investing, leads to a 12.1% increase in monthly trading volume. Estimated in a separate set of firms, Panel B Column 1 shows that switching from the Russell 2000 to the Russell 1000, which causes a decrease in passive investing, leads to a 9.3% decrease in monthly trading volume. These findings suggest that higher index fund holdings lead to higher trading volume. Column 2 shows that index switching leading to an increase (decrease) in passive investing leads to a 1.9% increase (1.5% decrease) in short interest. This finding is intuitive, since a significant part of many passive funds' revenue stream is derived from lending the shares that they hold to short-sellers.[20]

We next examine the effects of index investing on return volatility and stock price corre-

[20]This result is consistent with the findings in Palia and Sokolinski (2019) who examine the relation between index investing and short selling in greater depth. They show an increase in passive investing leads to more securities lending which, in turn, leads to more short selling.

lations. Remark 1 in Section II.B.1 predicts that an increase in index investing should lead to more price pressure and hence more volatility and a change in asset correlations. Column 3 shows that an increase in index investing raises the idiosyncratic volatility of the individual stock, and a decrease in index investing lowers it. Both these results are consistent with those of Ben-David et al. (2018), and represent an out-of-sample validation of their findings, since their sample ends in 2006 and our sample begins in 2007.[21] Columns 4 and 5 show that an increase in passive investing raises the stock's correlation with *both* the small-cap and large-cap indexes, and vice versa.[22] Column 6 shows that an increase in passive investing raises the stock's market beta, and vice versa. In other words, passive investing raises index assets' correlation with the market as a whole, consistent with the theoretical predictions of Basak and Pavlova (2013) and Baruch and Zhang (2019) and the price pressure prediction in Section II.B.1.[23]

E. Effects on Information Production

The results so far show that index investing does change certain aspects of financial markets. In particular, it leads to a change in the mass of investors in individual stocks, as well as a change in correlations and increased idiosyncratic return volatility. Remark 3 in Section II.B.1 predicts that a decrease in the mass of active investors will lead to a reduction in information production. Accordingly, we next examine the effects of index investing on information production. We examine three measures of information-gathering by investors. First, we use the Google Search Volume Index (SVI) for each stock and day. This measure is an index of internet search volume for the stock and is a commonly used measure of attention by retail investors (Engelberg & Gao, 2011). Second, we use the daily

[21]The findings are also consistent with an earlier literature documenting price pressure from index rebalancing (e.g., Shleifer (1986), Kaul, Mehrotra, and Morck (2000)).

[22]These findings are consistent with earlier work by Goetzmann and Massa (2003) and Da and Shive (2018), who document a correlation between index membership and return comovement.

[23]As noted in Wurgler (2011), this increase in beta is economically important as it may generate real economic effects – changes to beta will affect a firm's cost of capital, thereby affecting corporate investment decisions.

number of page requests for each sample stock's filings on the EDGAR website computed from the Securities and Exchange Commission (SEC) EDGAR log files. This measure is an index of attention to the firm's fundamental information and disclosures via their EDGAR documents (Iliev, Kalodimos, & Lowry, 2020).[24] Finally, we compute the number of analyst reports about each sample firm, from the I/B/E/S dataset. This measure is an index of information production by professional equity analysts.

Table IV presents the results. In Panel A we see that moving from the Russell 1000 to the Russell 2000, which caused an increase in ownership by index funds, leads to a lower level of information production in all three measures. Specifically, we find that Google SVI falls by 3.8 log points, EDGAR page requests fall by 24.8 log points, and the number of analyst reports falls by 10.8 log points. These economic magnitudes are significant: For example, the 24.8 log point decrease relative to the baseline average of 5.32 (in logs) corresponds to moving from 205 to 159 EDGAR page requests per day (a decrease of 22%). The magnitude of this effect is similar to the change in passive fund ownership that we document – a 1.5 to 1.7 percentage point change in passive fund ownership corresponds to an approximate change of 15% in the level of pre-treatment ownership. However, there is considerable noise in these measures, and the result on Google SVI is statistically insignificant while the estimates on EDGAR page views and analyst reports are statistically significant at $p < 0.10$.

While the results in Panel A provide suggestive evidence that index investing leads to a reduction in information production, the lack of statistical significance also means that those estimates are noisy. Panel B shows that moving from the Russell 2000 to the Russell

[24]Our measure is not identical to the measure in Iliev et al. (2020). Iliev et al. (2020) map EDGAR requests for a firm's annual proxy statement to fund IP addresses in the period prior to each annual shareholder meeting. By contrast, we measure the daily number of requests for any of the firm's EDGAR filings by anyone over our entire sample period. Using the data from Iliev et al. (2020), which the authors generously shared with us, we obtain qualitatively similar results. In particular, we find firms that switch into the Russell 1000, leading to an increase in active ownership, experience an increase in downloads. Similarly, firms that switch into the Russell 2000, leading to an increase in passive ownership, experience a decrease in downloads. However, because the sample in Iliev et al. (2020) looks at EDGAR requests for the firms proxy statement only in the period immediately prior to each annual shareholder meeting, there are only 110 and 74 observations, respectively, in our sample for the upper and lower bounds. As such, this test has poor power and the results are not statistically significant at the usual levels.

1000, which caused a drop in ownership by index funds, produced a higher level of information production in all three measures. Here, some of the coefficients are more precisely estimated because we have nearly twice as many stocks in this subsample. As a result, we find statistically significant results for EDGAR downloads and analyst reports. Moreover, the estimates are strikingly symmetric with the lower-band estimates: a lower level of index investing upon crossing the upper band is followed by a 4.4 log point increase in Google SVI, a 20.3 log point increase in EDGAR page views, and a 14.3 log point increase in the number of analyst reports.

In short, the results in this section are consistent with the predictions of the augmented Grossman and Stiglitz (1980) model. Specifically, because index investing leads to a decrease in the mass of active investors, but does not change the relative incentive among active investors to gather information, there is less information production about individual stocks.[25] To the best of our knowledge, our paper is the first to show evidence on this question.

F. Effects on Price Informativeness

A number of theoretical models examine the relation between index investing and price efficiency. Notably, Remark 4 of the augmented Grossman and Stiglitz (1980) model predicts a non-result: in equilibrium, the fraction of active investors who pay to become privately-informed should adjust to a change in the composition of investors such that price informativeness is unchanged. Accordingly, we next examine the effects of passive investing on the information content of stock prices. First, we examine weak-form price efficiency. We compute variance ratios for each month of sample stocks' returns over horizons of $q = 4$ and 8 trading days. Formally, we use the q-period bias-corrected variance ratio test of Lo and

[25]Our model and results are also consistent with the findings of Iliev et al. (2020) who show that index funds gather less information about their portfolio firms than active funds.

MacKinlay (1988):

$$VarRatio(q) = \frac{\hat{\sigma}^2(q)}{q \times \hat{\sigma}^2}, \tag{8}$$

where

$$\hat{\sigma}^2(q) = \frac{k}{(n-q+1)(k-1)} \sum_{t=q}^{n} (p_t - p_{t-q} - q\hat{\mu})^2, \tag{9}$$

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{t=1}^{n} (p_t - p_{t-1} - \hat{\mu})^2, \tag{10}$$

$$\hat{\mu} = \frac{1}{n} \sum_{t=1}^{n} (p_t - p_{t-1}), \tag{11}$$

and the data consists of $kq+1$ observations (for convenience we define $n = kq$). We calculate variance ratios separately for each firm and year using overlapping observations within the month. Formally, we examine the absolute value of the centered ratio:

$$AbsVarRatio_t(q) = abs(VarRatio_t(q) - 1) \tag{12}$$

The efficient benchmark is a variance ratio equal to 1 – that is, returns over a q-day horizon had a variance that was q times the variance of daily returns. Thus, $AbsVarRatio = 0$ defines perfect weak-form efficiency, and the larger the value of $AbsVarRatio$, the further the stock price process is from the random-walk benchmark as in Boehmer and Kelley (2009). In other words, we measure how far the stock price process deviated from a pure random walk.

Table V columns 1 and 2 present our difference-in-differences estimates around the Russell index discontinuities. We find that changes in passive investing did not lead to any significant changes in absolute variance ratios over either the 4 or 8 day horizon. That is, even though index investing changes the composition of investors, it does not have any significant effect on stocks' weak-form price efficiency.

It is important to note that our finding of no effect is not due to a lack of statistical power. For example, in Panel A (stocks around the lower yearly Russell band) the estimated treatment effect from index assignment is a change of -0.017 to the 4-day absolute variance ratio. How large is this change? We show the MDES below the estimate; the MDES shows that the standard deviation of the 4-day AVR in the lower-band sample is 0.413, so the estimated effect is less than one-twentieth of one standard deviation. Moreover, the minimum detectable effect size (MDES) for the AVR in this estimate is +/-0.068, so our research design has the ability to reliably detect a treatment effect on the order of (0.068 / 0.413) = one-sixth of one standard deviation. Yet we find no effect – precisely estimated. The results are similar for both the 8-day AVR and in the upper-band subsample.

Next we investigate stock-level mispricing. To do this, we use the mispricing measure of Stambaugh et al. (2015), which measures the combined ranking of each stock on 11 documented asset-pricing anomalies each month. The average mispricing level (*misp*) for our sample of stocks is 0.501 compared to a level of 0.503 across the entire universe of CRSP stocks used by Stambaugh et al. (2015). Thus, our sample of mid-cap stocks near the Russell thresholds appears to be representative of the broader universe of stocks in terms of mispricing.

Table V Column 3 shows that in both subsamples, a change in passive investing in the stock was not associated with any significant change in the anomaly-mispricing score of the stock. Again, we note that the insignificant effects on *Misprice* are unlikely a result of insufficient power. We see from the MDES that our research design has enough statistical power to detect a treatment effect on the order of one-third of one standard deviation in the lower-band sample, and one-fifth of one standard deviation in the upper-band sample – five percentile ranks and three percentile ranks on a range of 1-100, respectively. Yet in both cases we find no effect. Thus, using the mispricing measure of Stambaugh et al. (2015) we again find a zero treatment effect that is precisely estimated, indicating that index investment does not significantly change treated stocks' price efficiency.

Finally, we investigate the incorporation of information into stock prices. To do this, we use a simple and well-documented source of return predictability that varies by stock: post-earnings-announcement drift (PEAD). Our measure is based on earnings announcements from 1 to 24 months post-index assignment for each sample stock-year. Specifically, for an earnings announcement on day t, we regress the cumulative stock return from day $t+3$ to day $t+63$ on the standardized unexpected earnings (SUE) of the announcement, which is scaled so a value of 1 represents a surprise in earnings per share equal to 1% of the stock price. The coefficient of returns on SUE, for each stock-year in the sample, is denoted $Beta_{PEAD}$. A value of $Beta_{PEAD} = 0$ corresponds to no post-earnings-announcement drift i.e. the stock incorporates all the information contained in the earnings announcement, on average, within 2 trading days after the announcement.

The mean $Beta_{PEAD}$ in our sample is 4.3%; in other words, a 1% standard deviation positive surprise in the stock's earnings announcement is followed by a predictable post-announcement return drift of 4.3%. The sample mean has a standard error of 0.6% and so the mean is strongly different from zero; that is, in our sample of mid-cap stocks, there is significant PEAD on average. However, the sample stocks are also quite heterogeneous in their level of PEAD: the standard deviation of $Beta_{PEAD}$ is 17%.

Table V Column 4 shows that passive investing is not associated with any change in the speed of incorporation of new information into stock prices. Once again, both the estimated treatment effect and the minimum detectable effect size (MDES) are significantly smaller than one standard deviation in the outcome variable in both subsamples, so the finding of zero effect is not a type II error (i.e. it is not due to a lack of power).

Taken together, our results on liquidity, variance ratios, mispricing and post-earnings-announcement drift all suggest that index investing does not alter price efficiency or market function. We stress that these non-results are informative given the precision in our estimated coefficients. Indeed, Abadie (2018) notes that *"...rejection of a point null often carries very little information, while failure to reject may be highly informative."* Accordingly, the non-results in this section provide strong evidence regarding the impact of index investing on the informativeness of stock prices.

V. Conclusion

We examine the effects of the rise in index investing. If index investors are passive in the "buy and hold" sense then they are, by definition, free-riding off the information production of active managers. Yet, to date, it is unclear whether this shift in investor composition affects price informativeness. Put differently, not everyone can index, some investors have to be *active* managers. The question is, how many active managers are enough to ensure that prices correctly reflect fundamental values?

We augment the Grossman and Stiglitz (1980) model to include a choice by investors – investors choose to be passive, active and publicly-informed, or active and privately-informed. While several existing models predict that a rise in index investing will alter the information in prices, our augmented model predicts no effect. We then test between these two different classes of models. Using Russell Index reconstitutions as a source of exogenous variation, we find that index investing changes the composition of investors and it affects trading dynamics. An exogenous increase in index fund ownership leads to less ownership by active funds, higher turnover, increased correlations with other members of the Russell 2000 index, and increased short-selling. We then explore whether index investing affects information production and/or price efficiency.

Consistent with the predictions in the augmented Grossman and Stiglitz (1980) model, we find that an increase in the mass of passive investors leads to lower information production at the firm-level. Nevertheless, when we examine measures of price informativeness, we find no effects. Treated stocks do not experience any difference in variance ratio tests, post-earnings announcement drift, or anomaly mispricing. The results confirm an important, but untested, prediction from Grossman and Stiglitz (1980): Following an exogenous change in investor composition, we find evidence that the mass of active informed investors adjusts such that price informativeness is unchanged.

We are careful to note that both our theoretical and empirical results rely on assumptions. Theoretically, we assume investors can be categorized into one of three groups. In practice, investors may not fall perfectly into one category (e.g., Robertson (2019, 2019)). Moreover, as noted by Pedersen (2018), active funds do not have to underperform passive funds after fees because real-world rebalancing frictions can affect the performance of both passive and active funds.

Empirically, we caution that the external validity of our results remains unknown. The difference-in-differences methodology we employ estimates the local average treatment effect (LATE) of index assignment around the upper and lower Russell bands both narrow slices of mid-cap U.S. stocks. As such, it is not clear whether our estimates generalize to larger changes or higher levels of index investing. Moreover, if index switching causes other changes that are not due to changes in ownership, this could confound our inferences (for example, stocks at the bottom of the R1000 may be less important in their index, since they have a lower weight, which could lead to lower attention and/or lower turnover).

However, our evidence suggests that index reconstitutions lead to a large change in ownership, suggesting this is the primary channel through which index reconstitutions affect firms. We also discipline our tests by first writing a model that makes the same predictions at nearly all levels of index investing, and our empirical results are consistent with those predictions. In particular, the models sharpest prediction is that there is *no* effect on price informativeness, which is what we find in the data. Zooming out from the Russell subsamples to examine all U.S. equities, Figure 5 shows a similar pattern in the aggregate time series: As passive investing increased rapidly from 2007-2016, there was no significant change in average price efficiency across all U.S. listed stocks.

Thus, notwithstanding these caveats, we believe our estimates provide important data about an important question. More broadly, we note that our model's theoretical predictions hold for *any* feasible level of passive investing, which suggests our estimates would hold for larger or smaller changes in index investing. In addition, the overall trends in index fund ownership and price efficiency in recent years are consistent with our predictions and our cleanly identified estimates. Figure 5 plots the yearly average levels of passive ownership and inverse price efficiency (measured by the 4-day variance ratio) across all Russell 3000 stocks during our sample period from 2007 to 2016. In this ten year period, index fund ownership quintupled from 2% to 11% of market capitalization on average. Yet over the same period, the average variance ratio *fell* slightly from 1.16 to 1.10, reflecting a slight *improvement* in price efficiency. The conclusions are similar when we weight the averages by market capitalization. Overall, all of our analyses: the theoretical predictions, difference-in-differences estimates, and broad sample trends, point to the same conclusion.

Our findings relate to the recent controversy in the investment community about whether passive investing damages social welfare. To the extent that equity markets help allocate capital in the economy, some have argued that index investing damages economic efficiency. For example, Fraser-Jenkins et al. (2016) argue that an economy with substantial passive investing is worse than a centrally-planned economy. In response to this and similar assertions, Libson and Parchomovsky (2020) argue that regulators should tax passive investment to help defray the costs of active management. While we are careful to note that our empirical evidence cannot provide welfare implications, our results suggest these concerns might be unfounded. Our results show that index investing changes the composition of investors, however price informativeness is unchanged. Accordingly, our findings imply that frictions that inhibit trading by arbitragers or limit the realignment of active traders in the market appear to be minimal. Of course, our analysis does not say whether the social and individual margins for being active versus passive differ. We leave this important question for future research.

References

Abadie, A. (2018). Statistical non-significance in empirical economics. *NBER Working Paper*. Retrieved from http://www.nber.org/papers/w24403.pdf

Appel, I., Gormley, T. A., & Keim, D. B. (2016). Passive investors, not passive owners. *Journal of Financial Economics*, *forthcoming*.

Appel, I., Gormley, T. A., & Keim, D. B. (2020). Identification using russell 1000/2000 index assignments: A discussion of methodologies. *Working Paper*.

Asparouhova, E., Bessembinder, H., & Kalcheva, I. (2013). Noisy prices and inference regarding returns. *The Journal of Finance*, *68*(2), 665–714.

Baltussen, G., van Bekkum, S., & Da, Z. (2019). Indexing and stock market serial dependence around the world. *Journal of Financial Economics*, *132*(1), 26–48.

Baruch, S., & Zhang, X. (2019). The distortion in prices due to passive investing. *Working Paper*.

Basak, S., & Pavlova, A. (2013). Asset prices and institutional investors. *American Economic Review*, *103*, 1728-1758.

Ben-David, I., Franzoni, F., & Moussawi, R. (2018). Do etfs increase volatility? *The Journal of Finance*, *73*(6), 2471–2535.

Ben-David, I., Franzoni, F., & Moussawi, R. (2019). An improved method to predict assignment of stocks into russell indices. *working paper*.

Bloom, H. S. (1995). Minimum detectable effects: A simple way to report the statistical power of experimental designs. *Evaluation Review*, *19*(5), 547-556.

Boehmer, E., & Kelley, E. K. (2009). Institutional investors and the informational efficiency of prices. *Review of Financial Studies*, *22*, 3563-3594.

Bogle, J. C. (2016). The index mutual fund: 40 years of growth, change, and challenge. *Financial Analyst Journal*, *January/February*, 9-13.

Bond, P., & Garcia, D. (2019). The equilibrium consequences of indexing. *Working Paper*.

Boone, A. L., & White, J. T. (2015). The effect of institutional owners on firm transparency and information production. *Journal of Financial Economics*, *117*, 508-533.

Breugem, M., & Buss, A. (2019). Institutional investors and information acquisition: Implications for asset prices and informational efficiency. *Review of Financial Studies*.

Brown, D. C., & Davies, S. W. (2017). Moral hazard in active asset management. *Journal of Financial Economics*, *1*, 1-15.

Budish, E., Crampton, P., & Shim, J. (2015). The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics*, *130*.

Chabakauri, G., & Rytchkov, O. (2021). Asset pricing with index investing. *Journal of Financial Economics*.

Chang, Y.-C., Hong, H., & Liskovich, I. (2015). Regression discontinuity and the price effects of stock market indexing. *Review of Financial Studies*, *28*, 212-246.

Cong, L. W., & Xu, D. X. (2016). Rise of factor investing: Asset prices, informational efficiency, and security design. *Working Paper*.

Crane, A. D., Michenaud, S., & Weston, J. P. (2016). The effect of institutional ownership on payout policy: Evidence from index thresholds. *Review of Financial Studies*, *29*, 1377-1408.

Cremers, M., Ferreira, M. A., Matos, P., & Starks, L. (2016). Indexing and active fund management: International evidence. *Journal of Financial Economics*, *120*(3), 539–560.

Da, Z., & Shive, S. (2018). Exchange-traded funds and asset return correlations. *European Financial Management*, *24*, 136-168.

Davila, E., & Parlatore, C. (2019). Trading costs and informational efficiency. *Journal of Finance, Forthcoming*.

Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, *66*(5), 1461–1499.

Fraser-Jenkins, I., Gait, P., Harmsworth, A., Diver, M., McCarthy, S., Stancikas, R., ... Hughes, M. (2016). The silent road to serfdom: Why passive investing is worse than marxism. *Fund Management Strategy, Alliance Bernstein*.

Glossner, S. (2021). Russell index reconstitutions, institutional investors, and corporate social responsibility. *Critical Finance Review*, *forthcoming*.

Glosten, L., Nallareddy, S., & Zhou, Y. (2016). ETF activity and informational efficiency of underlying securities. *Working Paper*.

Goetzmann, W. N., & Massa, M. (2003). Index funds and stock market growth. *Journal of Business*, *76*, 1–28.

Goldstein, I., & Yang, L. (2017). Commodity financialization: Risk sharing and price discovery in commodity futures markets. *Working Paper*.

Grossman, S. J., & Stiglitz, J. (1980). On the impossibility of informationally efficient markets. *American Economic Review*, *70*, 393-408.

Haddad, V., Huebner, P., & Loualiche, E. (2021). How competitive is the stock market? theory, evidence from portfolios, and implications for the rise of passive investing. *Working Paper*.

Heath, D., Macciocchi, D., Michaely, R., & Ringgenberg, M. C. (2021). Do index funds monitor? *Review of Financial Studies*, *forthcoming*.

Heath, D., Ringgenberg, M. C., Samadi, M., & Werner, I. (2019). Reusing natural experiments. *Working Paper*.

Iliev, P., Kalodimos, J., & Lowry, M. (2020). Investors' attention to corporate governance. *Review of Financial Studies, forthcoming*.

Israeli, D., Lee, C. M., & Sridharan, S. A. (2016). Is there a dark side to exchange trade funds (ETFs)? An information perspective. *Working Paper*.

Kaul, A., Mehrotra, V., & Morck, R. (2000). Demand curves for stocks do slope down: New evidence from an index weights adjustment. *Journal of Finance*, *55*, 893–912.

Libson, A., & Parchomovsky, G. (2020). Reversing the fortunes of active funds. *Faculty Scholarship at Penn Law*.

Liu, H., & Wang, Y. (2019). Index investing and price discovery. *Working Paper*.

Lo, A., & MacKinlay, C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*(1), 41-66.

Malkiel, B. G. (2013). Asset management fees and the growth of finance. *The Journal of Economic Perspectives*, *27*(2), 97–108.

Mullins, W. (2014). The governance impact of index funds: Evidence from regression discontinuity. *Working Paper*.

Norges Bank Investment Management. (2022). *Investment strategy.* Retrieved January 2022, from https://www.nbim.no/en/the-fund/how-we-invest/investment-strategy/

Palia, D., & Sokolinski, S. (2019). Does passive investing help relax short-sale constraints? *Working Paper*.

Pedersen, L. H. (2018). Sharpening the arithmetic of active management. *Financial Analysts Journal*, *74*, 21–36.

Robertson, A. (2019). Passive in name only: Delegated management and 'index' investing. *Yale Journal on Regulation*, *36*.

Robertson, A. (2020). The (mis)uses of the s&p 500. *Working Paper*.

Schmidt, C., & Fahlenbrach, R. (2017). Do exogenous changes in passive institutional ownership affect corporate governance and firm value? *Journal of Financial Economics*, *124*(2), 285–306.

Shleifer, A. (1986). Do demand curves for stocks slope down? *Journal of Finance*, *41*, 579-590.

Stambaugh, R. F., Yu, J., & Yuan, Y. (2015). Arbitrage asymmetry and the idiosyncratic volatility puzzle. *The Journal of Finance*, *70*(5), 1903-1948. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12286 doi: 10.1111/jofi.12286

Stein, J. C. (1987). Informational externalities of welfare-reducing speculation. *Journal of Political Economy*, *95*, 1123-1145.

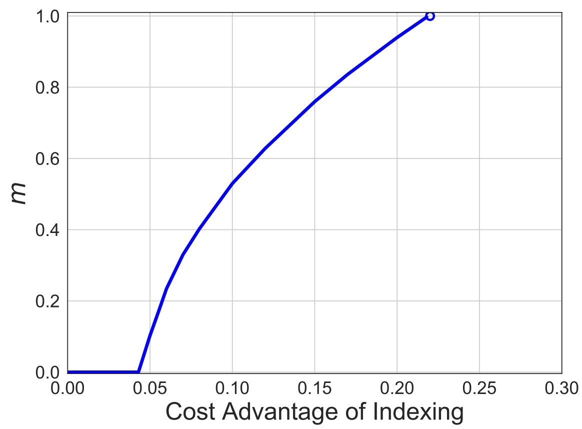
Subrahmanyam, A. (1991). A theory of trading in stock index futures. *The Review of Financial Studies*, *4*, 17-51.

Sunder, S. (1992). Market for information: Experimental evidence. *Econometrica: Journal of the Econometric Society*, 667–695.

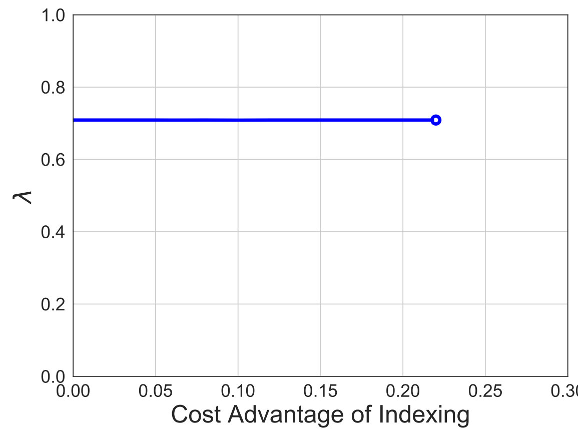
Wei, W., & Young, A. (2021). Selection Bias or Treatment Effect? A Re-Examination of

Russell 1000/2000 Index Reconstitution. *Critical Finance Review*, *forthcoming*.

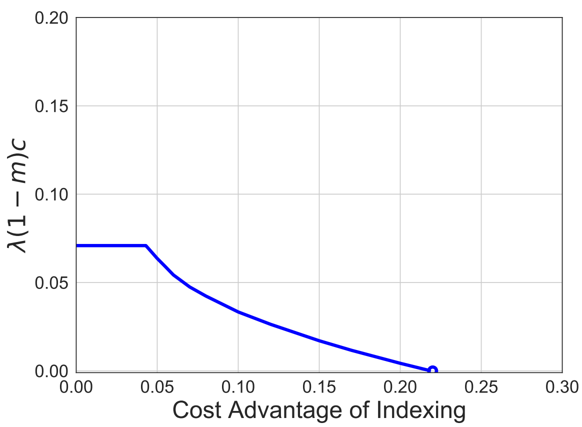
Wurgler, J. (2011). *On the economic consequences of index-linked investing.* Cambridge, MA: American Academy of Arts and Sciences.



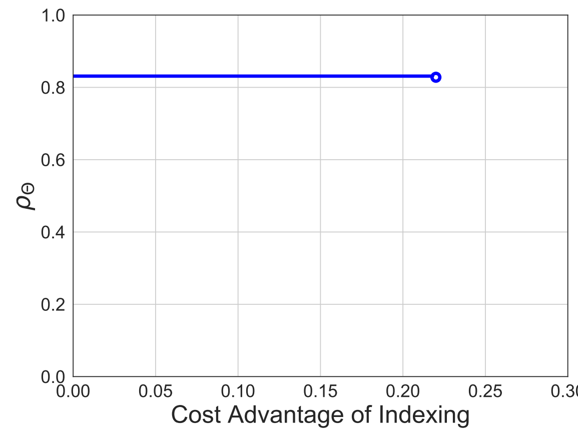
(a) Fraction of Investors who Index



(b) Fraction of Privately-Informed Active Investors



(c) Total Information Production



(d) Price Informativeness

Figure 1. Model Predictions

The figure plots comparative statics of the market equilibria in our version of the Grossman-Stiglitz (1980) model. The figures show changes in the market equilibrium as the cost advantage of index investing relative to active investing $(c_A - c_I)$ starts from zero and increases.

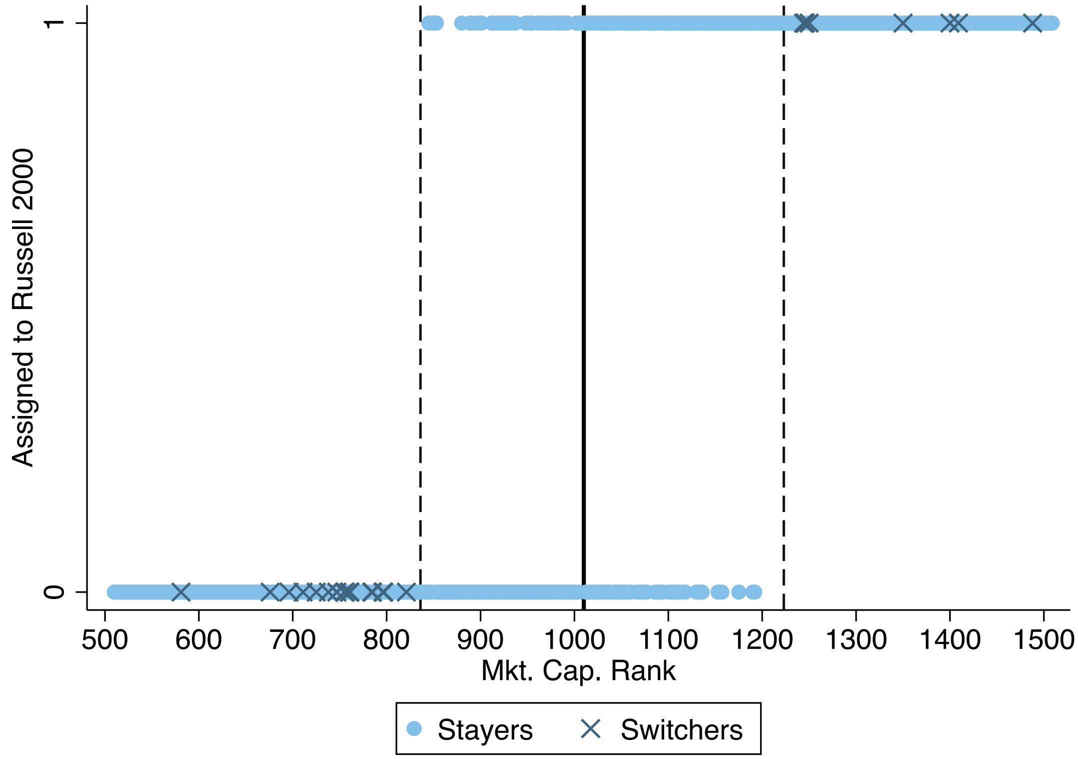


Figure 2. 2007 Russell Index Assignments

The figure plots stocks' assignments to the Russell 1000 and 2000 indexes in June of 2007 against our proxy for Russell's proprietary market cap rankings. Stocks near the rank-1000 threshold (solid line) all stayed in their previous index, breaking the discontinuity in index assignment. Close to the estimated upper and lower bands (dashed lines), however, there are sharp discontinuities in index switching.

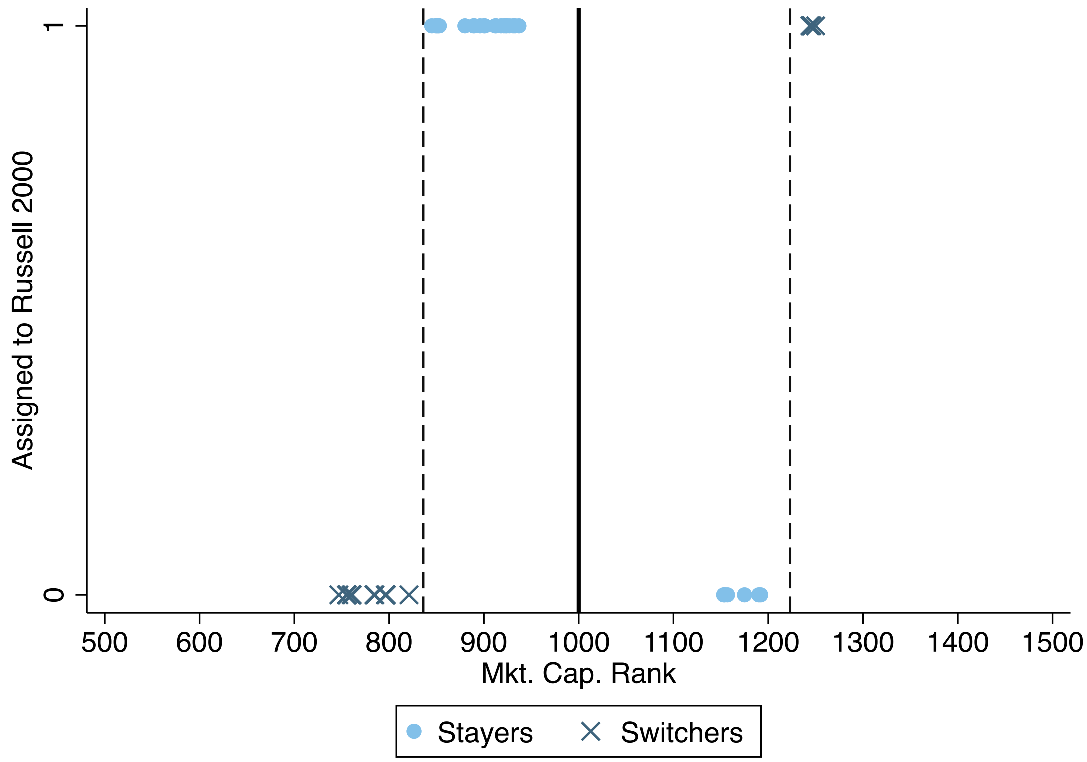
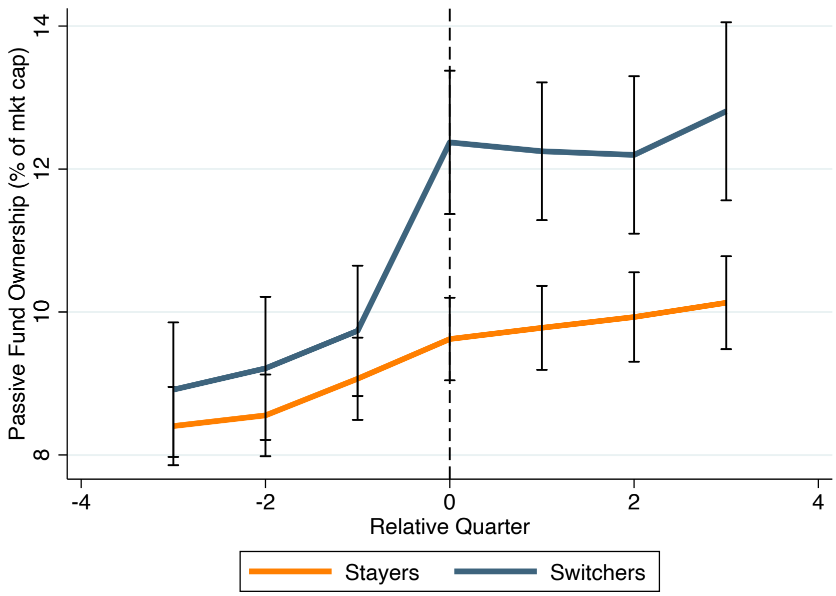
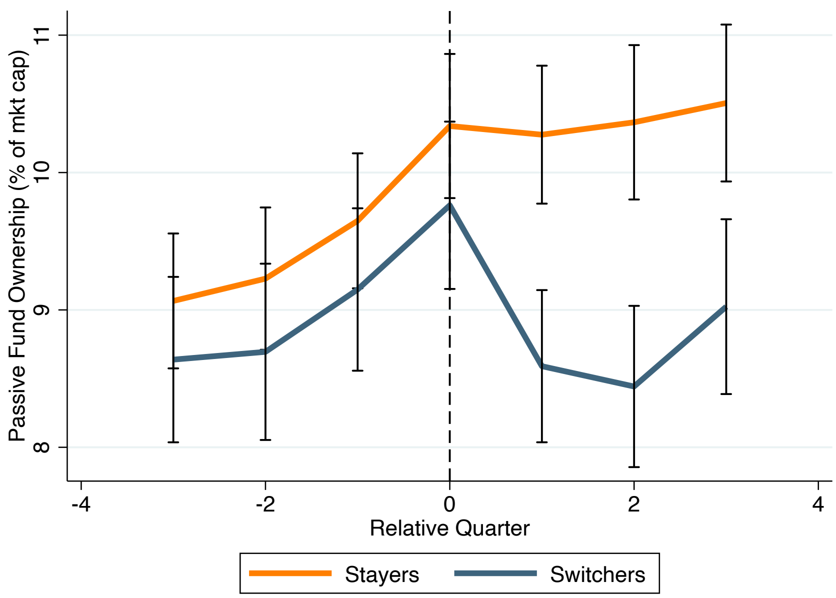


Figure 3. Sample Construction

The figure plots the sample stocks' assignment to the Russell 1000 and 2000 indexes in June of 2007 against our proxy for Russell's proprietary market cap rankings. The sample consists of stocks in a +/-100 rank around the upper and lower bands that had the potential to switch indexes from the previous year based on their lagged index membership.



(a) Lower Band



(b) Upper Band

Figure 4. Passive Ownership in Event Time

The figure plots the average ownership by passive funds of control (stayer) and treated (switcher) stocks by quarter, in event time relative to index assignment. The figure also shows 95% confidence intervals for each quarterly mean. The sample consists of observations for 3 quarters before and after index assignment, for stocks that were potential switchers within a +/-100 rank window of the yearly Russell bands from 2007-2016.

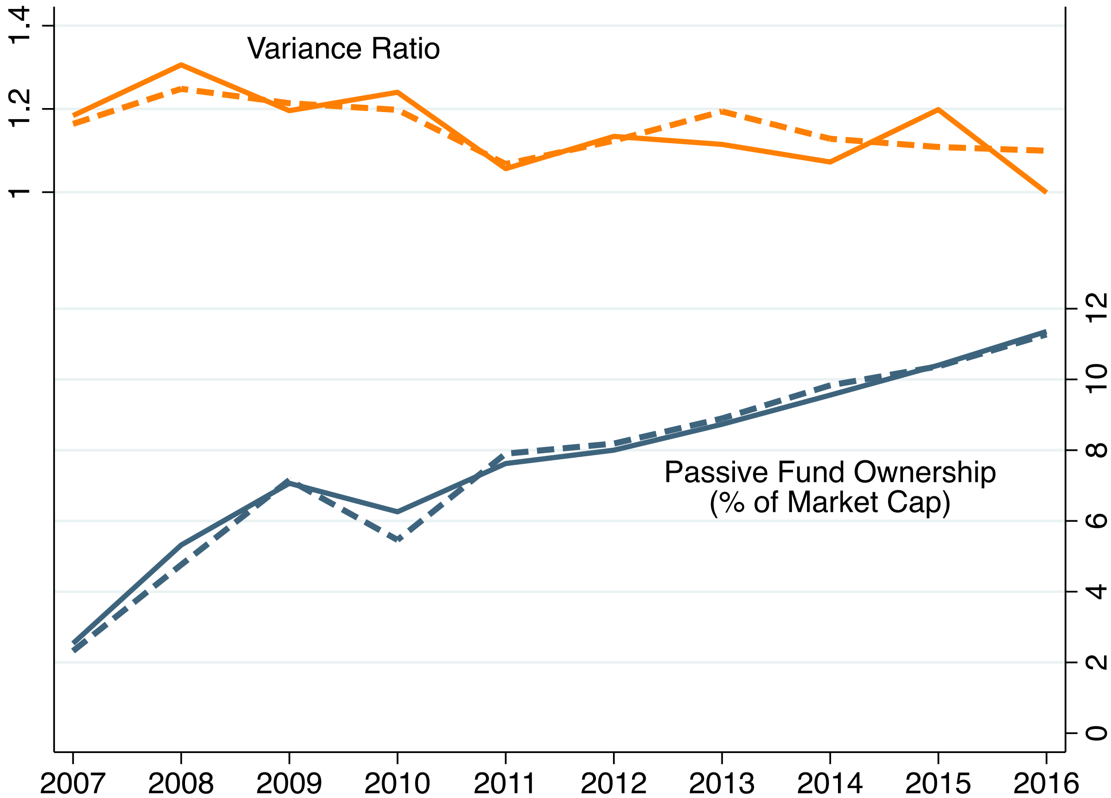


Figure 5. Overall Trends in Passive Ownership and Price Efficiency
The figure plots the average 4-day return variance ratio and average passive fund ownership across all stocks in the Russell 3000, yearly during our sample period from 2007-2016. Passive fund ownership is expressed as a percentage of each stock's market capitalization. Dashed lines indicate the equal-weighted average; solid lines indicate the value-weighted average.

Table I
Summary Statistics

The table presents summary statistics for ex ante stock characteristics: market capitalization, monthly trading turnover, monthly volatility, and categories of fund ownership. The sample consists of stocks that were potential switchers within a +/-100 rank window of the yearly Russell upper and lower bands from 2007-2016.

Panel A: Lower Band

	Mean	Std. Dev.	p10	p50	p90	N
Market Cap ($M)	1630.31	499.43	902.45	1647.34	2295.67	319
$FundOwn^{Passive}$	10.51	4.71	4.07	10.18	17.03	319
$FundOwn^{Active}$	23.19	11.56	7.78	23.42	37.25	319
$FundOwn$	33.70	13.45	15.33	33.59	50.63	319
$logTurn$	-1.22	0.76	-2.11	-1.17	-0.26	319
$LogVlt$	-3.83	0.52	-4.51	-3.83	-3.16	319
$ShortInt$	0.10	0.08	0.02	0.07	0.21	319
ρ^{R1000}	0.53	0.22	0.19	0.57	0.79	319
ρ^{R2000}	0.54	0.23	0.19	0.58	0.80	319
$MarketBeta$	1.26	0.70	0.51	1.21	2.16	319
$logSVI$	3.51	0.83	2.40	3.69	4.36	319
$logEDGAR$	5.32	1.56	3.76	5.66	6.86	319
$logAnalystReports$	2.14	0.84	0.69	2.30	2.94	319

Panel B: Upper Band

	Mean	Std. Dev.	p10	p50	p90	N
Market Cap ($M)	3019.27	875.14	1894.60	2946.14	4153.87	564
$FundOwn^{Passive}$	10.14	4.78	3.98	9.94	16.40	564
$FundOwn^{Active}$	25.62	12.48	8.93	25.86	40.00	564
$FundOwn$	35.76	14.43	16.54	36.43	52.60	564
$logTurn$	-1.34	0.68	-2.14	-1.28	-0.59	564
$LogVlt$	-3.96	0.49	-4.52	-3.93	-3.43	564
$ShortInt$	0.08	0.05	0.03	0.06	0.14	559
ρ^{R1000}	0.58	0.22	0.26	0.63	0.81	564
ρ^{R2000}	0.59	0.23	0.25	0.64	0.82	564
$MarketBeta$	1.23	0.62	0.50	1.24	1.99	564
$logSVI$	3.43	0.92	2.20	3.71	4.29	564
$logEDGAR$	4.99	1.75	3.25	5.44	6.64	564
$logAnalystReports$	2.12	0.76	1.10	2.30	2.83	564

Table II
Difference-in-differences Regression of Fund Ownership around Changes in Index Assignment

The table presents difference-in-differences estimates of the effects of Russell index switching on fund ownership, expressed as a percentage (1=1%) of the stock's market capitalization. The estimating equation is:

$$FundOwn_{it} = \beta\, IndexSwitch_{it} + \phi_i + \lambda_t + \epsilon_{it},$$

where $IndexSwitch_{it}$ equals one for stocks that switched indexes in quarters after the June Russell index reconstitution, and zero otherwise. $FundOwn^{R2000}$, $FundOwn^{R1000}$ denote ownership by passive funds that track the Russell 2000 and 1000 respectively. $FundOwn^{Passive}$ is total ownership by passive funds. $FundOwn^{Active}$ is total ownership by active funds. $FundOwn$ is total ownership by all funds. $InstOwn$ is total ownership by institutional investors that filed a 13F statement. The sample consists of quarterly observations for 3 quarters pre- and post- index reconstitution, for stocks that were potential switchers within a +/-100 rank window of the two Russell bands, each year from 2007-2016. Standard errors are robust and clustered by stock. ***: p<0.01, **: p<0.05, *: p<0.1.

Panel A: Lower Band

	(1) $FundOwn_{it}^{R2000}$	(2) $FundOwn_{it}^{R1000}$	(3) $FundOwn_{it}^{Passive}$	(4) $FundOwn_{it}^{Active}$	(5) $FundOwn_{it}$	(6) $InstOwn_{it}$
$R1000 \rightarrow R2000_i \times$	1.81***	-0.23***	1.81***	-0.45	1.36	5.37
$PostAssignment_t$	(26.5)	(-24.2)	(7.4)	(-0.6)	(1.6)	(1.2)
Observations	1,914	1,914	1,914	1,914	1,914	1,839
Adjusted R-squared	0.801	0.820	0.908	0.851	0.865	0.701
Stock × Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: Upper Band

	(1) $FundOwn_{it}^{R2000}$	(2) $FundOwn_{it}^{R1000}$	(3) $FundOwn_{it}^{Passive}$	(4) $FundOwn_{it}^{Active}$	(5) $FundOwn_{it}$	(6) $InstOwn_{it}$
$R2000 \rightarrow R1000_i \times$	-1.63***	0.23***	-1.16***	1.18*	0.01	-3.79***
$PostAssignment_t$	(-28.0)	(34.8)	(-5.5)	(1.8)	(0.0)	(-3.8)
Observations	3,384	3,384	3,384	3,384	3,384	3,237
Adjusted R-squared	0.817	0.844	0.880	0.837	0.833	0.860
Stock × Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes

Table III
Effects of Index Assignment on Asset Markets

The table presents difference-in-differences estimates of the effects of index assignment on the markets for individual stocks. The estimating equation is:

$$Y_{it} = \beta\, IndexSwitch_{it} + \phi_i + \lambda_t + \epsilon_{it},$$

where $IndexSwitch_{it}$ equals one for stocks that switched indexes in months after the June Russell index reconstitution, and zero otherwise. $LogTurn$ is the log of monthly volume divided by shares outstanding. $ShortInt$ is the mid-month short interest divided by shares outstanding. $LogVlt$ is the log of daily return volatility. ρ^{R1000} and ρ^{R2000} are the daily return correlation with the Russell 1000 and 2000 index respectively. $MarketBeta$ is the beta of the stock's daily returns on the daily CRSP value-weighted market return. MDES and Sample St.Dev. are the minimum detectable effect size and the sample standard deviation of each outcome variable. The sample consists of monthly observations for 11 months before and after index assignment, for stocks that were potential switchers within a +/-100 rank window of the two Russell bands each year from 2007-2016. Standard errors are robust and clustered by stock. ***: p<0.01, **: p<0.05, *: p<0.1.

Panel A: Lower Band

	(1) $LogTurn_{it}$	(2) $ShortInt_{it}$	(3) $LogVlt_{it}$	(4) ρ_{it}^{R1000}	(5) ρ_{it}^{R2000}	(6) $MarketBeta_{it}$
$R1000 \rightarrow R2000_i \times$	0.121***	0.019***	0.064*	0.046***	0.069***	0.300***
$PostAssignment_t$	(2.8)	(3.2)	(1.7)	(2.9)	(4.5)	(4.5)
MDES	±0.125	±0.017	±0.105	±0.045	±0.044	±0.190
Sample St.Dev.	0.829	0.081	0.612	0.237	0.233	0.986
Observations	6,643	6,614	6,642	6,450	6,450	6,642
Adjusted R-squared	0.819	0.807	0.687	0.394	0.388	0.278
Stock × Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × Month FE	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: Upper Band

	(1) $LogTurn_{it}$	(2) $ShortInt_{it}$	(3) $LogVlt_{it}$	(4) ρ_{it}^{R1000}	(5) ρ_{it}^{R2000}	(6) $MarketBeta_{it}$
$R2000 \rightarrow R1000_i \times$	-0.093***	-0.015***	-0.049**	-0.035***	-0.053***	-0.175***
$PostAssignment_t$	(-2.9)	(-4.6)	(-2.1)	(-3.9)	(-5.7)	(-4.5)
MDES	±0.090	±0.009	±0.066	±0.025	±0.026	±0.109
Sample St.Dev.	0.725	0.056	0.568	0.226	0.224	0.748
Observations	11,670	11,530	11,669	11,302	11,302	11,669
Adjusted R-squared	0.791	0.784	0.665	0.398	0.390	0.289
Stock × Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × Month FE	Yes	Yes	Yes	Yes	Yes	Yes

Table IV
Effects of Index Assignment on Information Production

The table presents difference-in-differences estimates of the effects of index assignment on measures of information production about individual stocks. The estimating equation is:

$$Y_{it} = \beta\, IndexSwitch_{it} + \phi_i + \lambda_t + \epsilon_{it},$$

where $IndexSwitch_{it}$ equals one for stocks that switched indexes in months (quarters) after the June Russell index reconstitution, and zero otherwise. $log(SVI)$ is the log of the monthly average Google Search Volume Index for the stock's ticker. $log(EDGAR)$ is the monthly average number of page requests per day for the firm's filings on the SEC EDGAR website. $log(AnalystReports)$ is the log of the quarterly number of nonduplicate equity analyst reports about the firm recorded in I/B/E/S. MDES and Sample St.Dev. are the minimum detectable effect size and the sample standard deviation of each outcome variable. For the first two measures the sample consists of monthly observations for 11 months before and after index assignment, for stocks that were potential switchers within a +/-100 rank window of the two Russell bands each year from 2007-2016. For the third measure the sample consists of quarterly observations for 3 quarters before and after index assignment. Standard errors are robust and clustered by stock. ***: p<0.01, **: p<0.05, *: p<0.1.

Panel A: Lower Band

	(1) $log(SVI)_{it}$	(2) $log(EDGAR)_{it}$	(3) $log(AnalystReports)_{it}$
$R1000 \rightarrow R2000_i \times$	-0.038	-0.248*	-0.108*
$PostAssignment_t$	(-1.0)	(-1.9)	(-1.8)
MDES	±0.107	±0.374	±0.166
Sample St.Dev.	0.822	1.880	0.853
Observations	6,699	6,699	1,914
Adjusted R-squared	0.836	0.787	0.717
Stock × Cohort FE	Yes	Yes	Yes
Year × Month FE	Yes	Yes	No
Year × Quarter FE	No	No	Yes

Panel B: Upper Band

	(1) $log(SVI)_{it}$	(2) $log(EDGAR)_{it}$	(3) $log(AnalystReports)_{it}$
$R2000 \rightarrow R1000_i \times$	0.044	0.203**	0.143***
$PostAssignment_t$	(1.5)	(2.5)	(2.8)
MDES	±0.085	±0.229	±0.144
Sample St.Dev.	0.897	1.962	0.826
Observations	11,844	11,844	3,384
Adjusted R-squared	0.830	0.809	0.598
Stock × Cohort FE	Yes	Yes	Yes
Year × Month FE	Yes	Yes	No
Year × Quarter FE	No	No	Yes

Table V
Effects of Index Assignment on Price Informativeness

The table presents difference-in-differences estimates of the effects of index assignment on measures of price informativeness. The estimating equation is:

$$Y_{it} = \beta\, IndexSwitch_{it} + \phi_i + \lambda_t + \epsilon_{it},$$

where $IndexSwitch_{it}$ equals one for stocks that switched indexes in months (quarters) after the June Russell index reconstitution, and zero otherwise. $AbsVarRatio^4$ and $AbsVarRatio^8$ are the 4-day and 8-day absolute variance ratios of price changes. $Misprice$ is the mispricing measure of Stambaugh et al. (2015). β^{PEAD} is the coefficient of the stock return over the three days around earnings announcements, regressed on the standardized unexpected earnings (SUE). MDES and Sample St.Dev. are the minimum detectable effect size and the sample standard deviation of each outcome variable. For the first three measures the sample consists of monthly observations for 11 months before and after index assignment, for stocks that were potential switchers within a +/-100 rank window of the two Russell bands each year from 2007-2016. For the last measure β^{PEAD} the sample runs for two years pre- and two years post-index assignment. Standard errors are robust and clustered by stock. ***: p<0.01, **: p<0.05, *: p<0.1.

Panel A: Lower Band

	(1) $AbsVarRatio^4_{it}$	(2) $AbsVarRatio^8_{it}$	(3) $Misprice_{it}$	(4) β^{PEAD}_{it}
$R1000 \rightarrow R2000_i \times$	-0.017	-0.027	0.001	-0.034
$PostAssignment_t$	(-0.7)	(-0.4)	(0.0)	(-1.0)
MDES	±0.068	±0.221	±0.048	±0.099
Sample St.Dev.	0.413	1.353	0.137	0.102
Observations	6,800	6,800	5,601	434
Adjusted R-squared	0.067	0.102	0.734	0.235
Stock × Cohort FE	Yes	Yes	Yes	Yes
Year × Month FE	Yes	Yes	Yes	No
Year FE	No	No	No	Yes

Panel B: Upper Band

	(1) $AbsVarRatio^4_{it}$	(2) $AbsVarRatio^8_{it}$	(3) $Misprice_{it}$	(4) β^{PEAD}_{it}
$R2000 \rightarrow R1000_i \times$	0.011	0.004	0.005	-0.014
$PostAssignment_t$	(0.7)	(0.1)	(1.1)	(-0.5)
MDES	±0.041	±0.132	±0.026	±0.079
Sample St.Dev.	0.385	1.246	0.126	0.185
Observations	11,928	11,927	9,905	845
Adjusted R-squared	0.062	0.095	0.739	0.223
Stock × Cohort FE	Yes	Yes	Yes	Yes
Year × Month FE	Yes	Yes	Yes	No
Year FE	No	No	No	Yes

Online Appendix for "On Index Investing"[1]

This appendix provides additional empirical and theoretical evidence to supplement the main text. Appendix 1.A provides additional empirical evidence that our treatment and control firms are as good as randomly assigned. Appendix 2.A provides additional theoretical detail on solving for the market equilibrium in our setting. Finally, Appendix 2.B shows the parameter choices used to numerically solve the model.

[1]Citation format: Coles, Jeffrey L., Davidson Heath, and Matthew C. Ringgenberg, Online Appendix for "On Index Investing," 2020, Working Paper.

Appendix 1. Supplemental Empirical Evidence

A. Balance Tests

First, we explore whether our treated and control samples are as good as randomly assigned across the upper and lower band. Although balance on covariates is not strictly necessary for unbiasedness in our setting (our difference-in-differences design requires parallel trends, not equal levels), the evidence below suggests our treatment and control groups across both bands are similar ex-ante.

Table A1 shows the results of two-sample comparison-of-means tests between switchers and stayers, on all the main outcome variables in the paper measured just prior to to treatment, across both bands. In every case the treated and control stocks appear similar *ex ante*. We conclude that our treated and control stocks appear similar on observables across both bands; the results suggest that selection bias is not a concern in our setting.

Table A1
Balance tests on *ex ante* observables

The table presents tests for potential selection bias in the sample by comparing *ex ante* stock characteristics across the groups of *ex post* index assignment, using the Russell index discontinuities. The sample consists of stocks that were potential switchers within a +/-100 rank window of the yearly Russell bands from 2007-2016. Standard errors are robust and clustered by stock. ***: p<0.01, **: p<0.05, *: p<0.1.

Panel A: Lower Band

	Stayers		Switchers		
	Mean	St.Dev.	Mean	St.Dev.	p-value
$FundOwn^{Passive}$	7.41	3.36	7.01	3.24	0.47
$Turnover$	1.38	0.53	1.52	0.60	0.17
$Volatility$	0.03	0.02	0.03	0.02	0.85
$AbsVarRatio^4$	0.37	0.29	0.48	0.45	0.13
$AbsVarRatio^8$	0.83	0.91	1.05	1.31	0.29
$Misprice$	0.55	0.14	0.54	0.13	0.60
β^{PEAD}	0.02	0.13	0.06	0.17	0.17

Panel B: Upper Band

	Stayers		Switchers		
	Mean	St.Dev.	Mean	St.Dev.	p-value
$FundOwn^{Passive}$	7.52	3.81	8.07	3.45	0.19
$Turnover$	1.34	0.52	1.38	0.40	0.39
$Volatility$	0.03	0.01	0.03	0.01	0.67
$AbsVarRatio^4$	0.37	0.35	0.37	0.38	0.92
$AbsVarRatio^8$	0.86	1.11	0.91	1.16	0.68
$Misprice$	0.48	0.12	0.49	0.14	0.57
β^{PEAD}	0.04	0.16	0.04	0.18	0.97

Appendix 2. Model Derivation

A. Market Equilibrium

We guess and check that the equilibrium price is linear in θ and x:

$$P = A + B\theta - Cx$$

A.1. Publicly-informed Active Investors

The publicly-informed active investor's prior about $\tilde{\theta}$ is N(0, σ_θ^2); they observe the signal

$$\frac{P + C\mu_x - A}{B} = \theta - \frac{C(x - \mu_x)}{B}$$

the residual of which has variance $\frac{C^2}{B^2}\sigma_x^2$.

The publicly-informed active investor's posterior about $\tilde{\theta}$ is:

$$E[\tilde{\theta}|P] = \frac{B^2\sigma_\theta^2}{C^2\sigma_x^2 + B^2\sigma_\theta^2}\frac{P + C\mu_x - A}{B} = \kappa\frac{P + C\mu_x - A}{B}$$

$$Var[\tilde{\theta}|P] = \frac{C^2\sigma_x^2}{C^2\sigma_x^2 + B^2\sigma_\theta^2}\sigma_\theta^2 = (1-\kappa)\sigma_\theta^2$$

, where $\kappa = \frac{B^2\sigma_\theta^2}{C^2\sigma_x^2 + B^2\sigma_\theta^2}$,

Publicly-informed active investors choose their position $\hat{y}$ by solving:

$$\max_y E\left[-exp(-\psi W_1)|P\right] \text{ subject to } W_1 = W_0 - c_A + y(\tilde{\theta} + \tilde{\epsilon} - P)$$

$$= E\left[-exp(-\psi(W_0 - c_A + y(\tilde{\theta} + \tilde{\epsilon} - P))|P\right]$$

$$= -exp\left(-\psi(W_0 - c_A + y(E(\tilde{\theta}|P) - P) + \frac{1}{2}\psi^2 y^2[Var(\tilde{\theta}|P) + \sigma_\epsilon^2]\right)$$

$$\frac{d(...)}{dy} = 0 = -exp\,(...) * \left(-\psi(E(\tilde{\theta}|P) - P) + \psi^2 y[Var(\tilde{\theta}|P) + \sigma_\epsilon^2]\right)$$

$$\hat{y}_U = \frac{1}{\psi}\frac{E(\tilde{\theta}|P) - P}{Var(\tilde{\theta}|P) + \sigma_\epsilon^2}$$

$$= \frac{1}{\psi}\frac{\kappa\frac{P + C\mu_x - A}{B} - P}{(1-\kappa)\sigma_\theta^2 + \sigma_\epsilon^2}$$

Their expected utility ex ante is:

$$E\left[-exp(-\psi(W_0 - c_A + \hat{y}_U(\tilde{\theta} + \tilde{\epsilon} - P)))\right]$$

A.2. Privately-informed Active Investors

Privately-informed active investors choose their position $\hat{y}$ by solving:

$$\max_y E\left[-exp(-\psi W_1)|\theta, P\right] \text{ subject to } W_1 = W_0 - (c_A + c) + y(\theta + \tilde{\epsilon} - P)$$

$$= E\left[-exp(-\psi(W_0 - (c_A + c) + y(\theta + \tilde{\epsilon} - P))\right]$$

$$= -exp\left(-\psi(W_0 - (c_A + c) + y(\theta - P)) + \frac{1}{2}\psi^2\sigma_\epsilon^2 y^2\right)$$

$$\frac{d(...)}{dy} = 0 = -exp\,(...) * (-\psi(\theta - P) + \psi^2\sigma_\epsilon^2 y)$$

$$\hat{y}_I = \frac{1}{\psi}\frac{\theta - P}{\sigma_\epsilon^2}$$

Their expected utility ex ante is:

$$E\left[-exp(-\psi(W_0 - (c_A + c) + \hat{y}_I(\theta + \tilde{\epsilon} - P))\right]$$

A.3. Index Investors

Index investors choose their position $\hat{y}$ by solving:

$$\max_y E\left[-exp(-\psi W_1)\right] \text{ subject to } W_1 = W_0 - c_I + y(\tilde{\theta} + \tilde{\epsilon} - \tilde{P})$$

$$= E\left[-exp(-\psi(W_0 - c_I + yE[(\tilde{\theta} + \tilde{\epsilon} - \tilde{P}))\right]$$

setting $\sigma^2_{Ind} = (1-B)^2\sigma^2_\theta + \sigma^2_\epsilon + C^2\sigma^2_x$:

$$= -exp(-\psi(W_0 - c_I + E[\tilde{\theta} - \tilde{P}] + \frac{1}{2}\psi^2 y^2 \sigma^2_{Ind})$$

$$\hat{y}_{Ind} = \frac{1}{\psi}\frac{E[\tilde{\theta} - \tilde{P}]}{\sigma^2_{Ind}} = \frac{1}{\psi}\frac{C\mu_x - A}{\sigma^2_{Ind}}$$

Their expected utility ex ante is:

$$= -exp(-\psi(W_0 - c_I + (C\mu_x - A)\frac{1}{\psi}\frac{C\mu_x - A}{\sigma^2_{Ind}}) + \frac{1}{2}\psi^2\Big(\frac{1}{\psi}\frac{C\mu_x - A}{\sigma^2_{Ind}}\Big)^2\sigma^2_{Ind})$$

A.4. Equilibrium Price Function

The equilibrium price function clears the market:

$$\tilde{x} = mX_{Ind} + \lambda(1-m)X_I + (1-\lambda)(1-m)X_U$$

$$\psi\frac{\tilde{x}}{1-m} = \frac{m}{1-m}\frac{C\mu_x - A}{\sigma^2_{Ind}} + \lambda\frac{\theta - P}{\sigma^2_\epsilon} + (1-\lambda)\frac{\frac{\kappa}{B}(P + C\mu_x - A) - P}{\sigma^2_U}$$

$$\sigma_U^2\psi\frac{\tilde{x}}{1-m} = \sigma_U^2\frac{m}{1-m}\frac{C\mu_x - A}{\sigma_{Ind}^2} + \frac{\lambda}{\sigma_\epsilon^2}\sigma_U^2\theta - \frac{\lambda}{\sigma_\epsilon^2}\sigma_U^2 P + (1-\lambda)(\frac{\kappa}{B}-1)P + (1-\lambda)\frac{\kappa}{B}(C\mu_x - A)$$

$$\frac{\lambda\sigma_U^2}{\sigma_\epsilon^2}P - (1-\lambda)(\frac{\kappa}{B}-1)P = [\frac{\sigma_U^2}{\sigma_{Ind}^2}\frac{m}{1-m} + (1-\lambda)\frac{\kappa}{B}](C\mu_x - A) + \frac{\lambda\sigma_U^2}{\sigma_\epsilon^2}\theta - \frac{\sigma_U^2\psi}{1-m}\tilde{x}$$

define $\xi = \left[\frac{\lambda\sigma_U^2}{\sigma_\epsilon^2} - (1-\lambda)(\frac{\kappa}{B}-1)\right]^{-1}$:

$$P = \xi[\frac{\sigma_U^2}{\sigma_{Ind}^2}\frac{m}{1-m} + (1-\lambda)\frac{\kappa}{B}](C\mu_x - A) + \frac{\xi\lambda\sigma_U^2}{\sigma_\epsilon^2}\theta - \frac{\xi\sigma_U^2\psi}{1-m}\tilde{x}$$

This shows that the linear price rule is an equilibrium, and pins down the values of A, B, C. In particular, the publicly-informed investor's signal is

$$w_\lambda = \theta - \frac{C(x-\mu_x)}{B} = \theta - \frac{\psi\sigma_\epsilon^2}{(1-m)\lambda}(x-\mu_x)$$

Further, because coefficient B does not change with m and coefficient C is increasing in m, the volatility of the asset price $\tilde{P}$ increases with the mass of index investors m, which generates our prediction 1.

B. Details of Numerical Solutions

We simulate $\tilde{\theta}, \tilde{\epsilon}, \tilde{x}$ using the parameters:

- $\sigma_\theta^2 = 0.2$
- $\sigma_\epsilon^2 = 1.0$
- $\sigma_x^2 = 0.5$
- $\mu_\theta = 0.0$

- $\mu_\epsilon = 0$
- $\mu_x = 1.0$
- $\psi = 0.3$
- $c = 0.1$

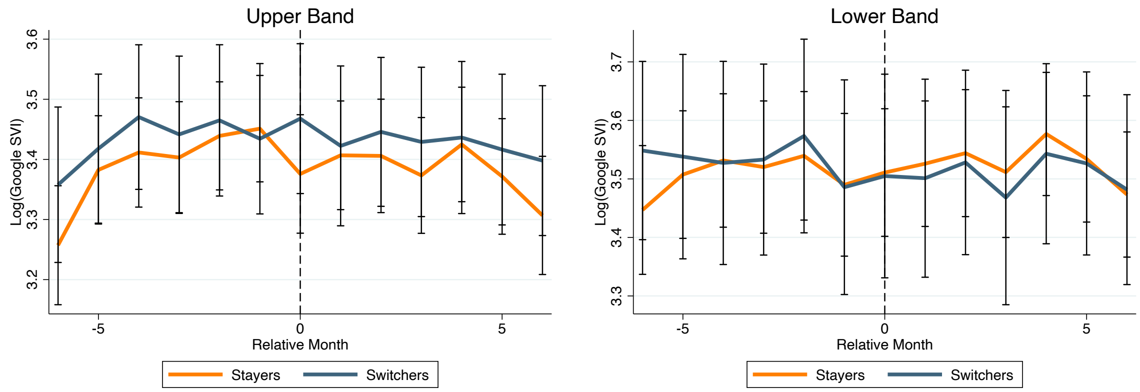
Given m and λ we solve the equilibrium price function and other market outcomes, then solve for m and λ that equalize expected utility across the three investor categories.

C. Event-Time Graphs of Information Production

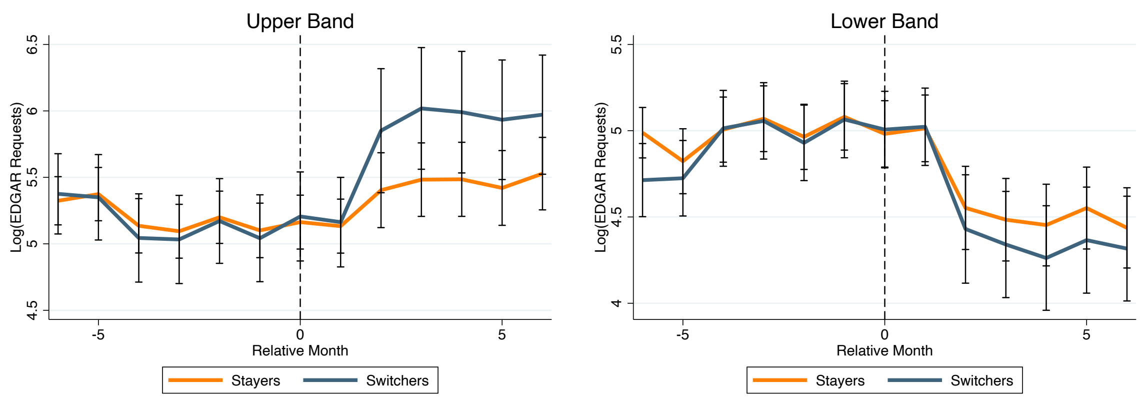
This section presents event-time graphs of the treatment effects of switching indexes on stock-specific information production. These graphs correspond closely to the estimates in Table IV.

Figure A1 plots event-time means for the two groups of sample stocks: Those that stayed in their index after reconstitution (stayers; control group) and those that switched indexes after reconstitution (switchers; treatment group). Panel A shows monthly means for Google Search Volume Index (SVI); Panel B shows monthly means for EDGAR page requests; Panel C shows quarterly means for the number of analyst reports per stock.

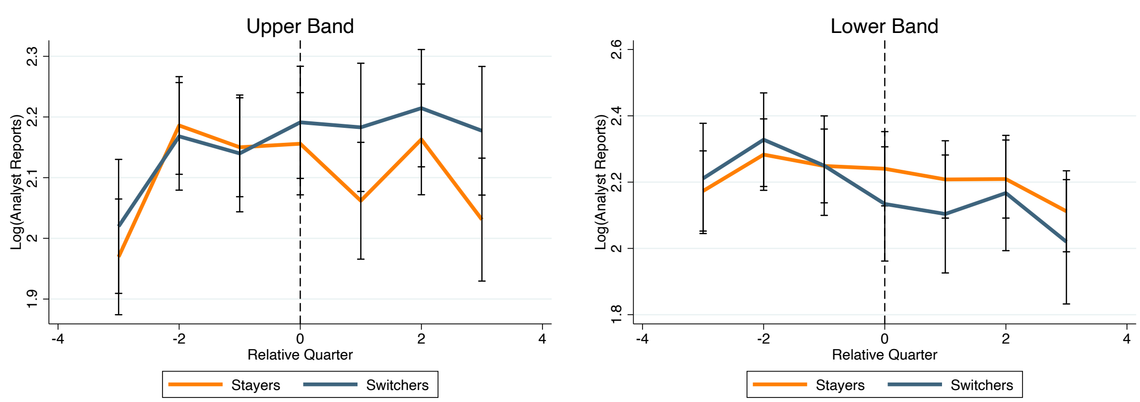
While the graphs are noisier than the event-time graphs for fund ownership (Figure 4), they support the parallel trends assumption. In all graphs, we see that index switchers and non-switchers appear similar prior to the reconstitution at t=0. For all three variables, we see that the line for switchers in the lower band (e.g., blue line in the left column graphs) is below the yellow line for non-switchers after t=0. Again, for all three variables, we see that the line for switchers in the upper band (e.g., blue line in the right column graphs) is higher than the yellow line for non-switchers after t=0.



(a) Google Search Volume Index



(b) EDGAR Page Requests



(c) Analyst Reports

Figure A1. The figure plots stock-level measures of information production for control (stayer) and treated (switcher) stocks, in event time relative to index assignment. The figure also shows 95% confidence intervals for each group mean in each period. The sample consists of stocks that were potential switchers within a +/-100 rank window of the yearly Russell bands from 2007-2016.