Complementarity of Passive and Active Investment on Stock Price Efficiency

Youngmin Choi*

16 November, 2017

Job Market Paper

Abstract

I investigate the *collective* impact of passive and active investment on stock price efficiency using a quasi-natural experiment. I document an improvement in efficiency due to an exogenous increase in passive investment, specifically in stocks widely held by actively managed funds. These active funds are compensated with higher realized returns after an exogenous increase in passive investment. I use the reconstitution of Russell indexes as an instrument. My findings suggest that active funds seek out inefficient stocks and ultimately experience superior returns due to the improvement in efficiency from passive investment. An increase in analyst following and a decrease in analyst forecast dispersion are identified as economic channels of the efficiency improvement. Overall, my results highlight the complementary role of passive and active investment on price discovery due to symbiotic nature of their existence.

*I am deeply indebted to my advisor Suzanne S. Lee for her invaluable guidance and support. I am also very grateful to Sudheer Chava, Alex Hsu, and Soohun Kim for their constant support and encouragement. I would like to thank Narasimhan Jegadeesh, Jeffrey Busse, Clifton Green, Rohan Ganduri, Peter Simasek, Nikhil Paradkar, and the seminar participants at Georgia Tech for their helpful comments and suggestions. I am responsible for any errors. Youngmin Choi is a doctoral student at the Scheller College of Business, Georgia Institute of Technology, 800 West Peachtree St NW, Atlanta, GA 30308; email: youngmin.choi@scheller.gatech.edu.

1 Introduction

Over the last 40 years, in which mutual funds have grown dramatically as investment vehicles,[1] passive and active investment have been at the center of many debates. Warren Buffett, who is regarded as one of the greatest stock market investors, in his annual letter stated "Over the years, I've often been asked for investment advice. ... My regular recommendation has been a low-cost S&P 500 index fund."[2] On the contrary, analysts at AllienceBernstein, a global asset management firm, in their report boldly titled "The Silent Road to Serfdom: Why Passive Investing Is Worse Than Marxism," argue forcefully that the rise of passive investing presents dangerous real-world barriers to the efficient allocation of capital in the economy.[3] While these two strategies are at opposite ends in terms of investment objective, each has been documented to have a significant role in influencing financial markets.[4] However, there is a lack of evidence as to how these two investment strategies, particularly jointly, affect the extent to which stock prices reflect all available information. In this paper, I examine the *collective* impact of passive and active investment on stock price efficiency using a causal setting.

The investment objectives of passively and actively managed funds reflect how they appreciate observed stock prices in terms of efficiency. Passive mutual funds aim to deliver the returns of a market index or benchmark portfolio (for example, S&P 500 index or Russell 2000 index) at a low cost because they consider stock prices to already be efficient. On the

[1]Since the inception of the first index fund, Vanguard 500 Index Fund, in 1976 with $11.3 million assets under management, the mutual fund industry has grown tremendously. At the end of 2016, the size of mutual fund industry in the U.S. is $16.3 trillion, according to the Investment Company Institute (see `http://www.icifactbook.org/ch2/17_fb_ch2`). This amount takes account for almost 80% of total market capitalization of the S&P 500 index as of December 2016.

[2] Source: `http://www.berkshirehathaway.com/letters/2016ltr.pdf`. Warren Buffett further recommends a S&P 500 index fund over hedge funds which are known as financial "elites" to provide absolute-returns using a long-short strategy.

[3]Source: `https://www.bloomberg.com/news/articles/2016-08-23/bernstein-passive-investing-is-worse-for-society-than-marxism`

[4]In regards to the impact of passive investment on corporate governance, see, for example, Appel, Gormley, and Keim (2016) and Boone and White (2015). Brav, Jiang, Partnoy, and Thomas (2008) document the impact of hedge funds on corporate governance. Among many others, Frazzini and Lamont (2008) document a relationship between mutual fund flows and the cross-section of stock returns. Cao, Liang, Lo, and Petrasek (2014) and Kokkonen and Suominen (2015) show that hedge funds improves the efficiency of stock prices.

contrary, active mutual funds aim to outperform a benchmark portfolio by selecting securities based on their research at a relatively higher cost because they believe observed prices of some stock do not fully reflect available information. By the nature of their investment objective, they need each other: passively managed funds need the presence of actively managed funds in order to have stock prices efficient enough, and actively managed funds also require a set of passively managed funds that gives active funds a comparative advantage of providing superior returns.

Due to the collaborative nature of the philosophy of passive and active investment, it is essential to examine the collective impact of the two types of investment on stock price efficiency. In the one extreme case, where all mutual funds are managed passively because stock prices are fully efficient, investors who spent effort and resources to gather and process information would not be compensated. This leads to the conclusion that a perfectly efficient market is impossible if information is costly (Grossman and Stiglitz (1980)). In the other extreme case if stock prices were sufficiently inefficient, many actively managed funds would be able to outperform their benchmarks. However, as more active funds exploit the inefficiencies, such opportunities become more elusive and prices become more efficient (Pástor and Stambaugh (2012)). The degree of inefficiency determines the willingness of investors to gather and trade on the information. Actively managed funds justify higher fees and expenses than passively managed funds as a compensation for their effort on security analysis (Ippolito (1989), Elton, Gruber, Das, and Hlavka (1993)).

In this paper, I analyze in two steps the question of how passive and active investment collectively affect stock price efficiency, and I document the complementary role of passive and active investment on stock price efficiency. In the first step, I separately investigate the impact of passive investment on stock price efficiency as no prior guidance exists to clearly show the directional association between passive investment and stock price efficiency.[5] My investigation reveals that an increase in passive investment improves stock price efficiency.

[5]I do not separately analyze the impact of active investment on stock price efficiency, as it is widely believed the positive impact of active investment on the efficiency. For active investment and stock price efficiency, see Cao, Liang, Lo, and Petrasek (2014) and Kokkonen and Suominen (2015) among many others.

In the second step, the collective impact of passive and active investment is examined by investigating how stock price efficiency varies based on the mix of passive and active investment in a stock. I find evidence that a material presence of active investors is necessary for passive investment to improve stock price efficiency. The finding implies the complementarity of passive and active investment in the efficient price discovery.

It is unclear how the investment of passive mutual funds affects price efficiency of securities. On the one hand, passive investment might inhibit price efficiency because passively managed funds would buy or sell a security depending on its relevance in mimicking a market or benchmarked index no matter how a security is priced relative to its fundamental (or intrinsic) value. In addition, because weights of individual stocks in passively managed funds are mechanically determined by their weights in the benchmarked portfolio or the index, the sensitivity of each stock to available information cannot be fully reflected in its weight in the portfolio. This mechanism of passive investment might lead to a breakdown of the link between the intrinsic value and the transaction prices of a security (Ben-David, Franzoni, and Moussawi (2014), Da and Shive (2016)). Recent debate in the finance industry argues that the current predominance of passive investment since the financial crisis in 2008 might undermine the efficient allocation of capital. On the other hand, passive investment might enhance the efficiency of securities as investors basically trade a basket of securities via passively managed funds, which enables information to be reflected for a broader set of stocks. Furthermore, stocks that experience an increase in passive investment exhibit an increase in liquidity and firm transparency (see, for example, Boone and White (2015)) and improvement in governance quality (see, for example, Appel, Gormley, and Keim (2016)).

Several empirical issues make it challenging to investigate the impact of passive investment on stock price efficiency. First and foremost, but often ignored in the asset pricing literature, a causal relationship between passive investment and stock price efficiency is subject to a severe endogeneity issue. That is, investment by passive mutual funds could be correlated with other factors, such as transaction cost or information asymmetry. For example, Falkenstein (1996) documents that mutual funds have a strong preference for stocks with low transaction costs, high liquidity, and low idiosyncratic volatility. Second, an ordinary least squares estimation of passive investment on stock price efficiency is subject to omitted variable bias. It is still questionable whether observable control variables and fixed effects can fully capture characteristics that simultaneously determine passive investment and stock price efficiency. To overcome these empirical challenges, I use the annual reconstitution of the Russell indexes to exploit exogenous variation in passive fund investment.

My identification strategy relies on two salient features of firms around the cutoff for Russell 1000 and 2000 indexes. First, firms on either side of the index threshold do not exhibit any systematic differences with respect to firm characteristics (For example, see Chang, Hong, and Liskovich (2015)). Second, because the Russell 1000 and 2000 indexes are value-weighted, firms near the top of the Russell 2000 index have significantly higher portfolio weights in the index compared with firms near the bottom of the Russell 1000 index. Thus, the value-weighted construction of the Russell indexes creates variation in passive fund investment that is plausibly exogenous to security price efficiency. These two characteristics of firms around the threshold allow us to exploit exogenous variation in passive investment.

The empirical design is based on two-stage least-squares specifications using the Russell indexes reconstitution as an instrument to overcome the endogeneity issue. The first stage examines passive fund holdings as a function of index inclusion at the threshold, and the second stage tests the impact of passive investment on price efficiency measures. In particular, in the first stage estimation, the empirical specification is a sharp regression discontinuity design to capture exogenous variation in passive investment around the Russell index threshold, similar to the specification used in Crane, Michenaud, and Weston (2016). In the second stage, I use the exogenous variation in passive fund holdings estimated from the first stage as an instrument to identify its impact on price efficiency. Using this empirical design, I find a stark difference in passive mutual fund holdings for stocks around the threshold of the Russell indexes. Investment of passively managed funds is about 33.4% higher for firms in the top 250 of the Russell 2000 index relative to firms in the bottom 250 of the Russell 1000 index. The difference is statistically and economically significant. However, I do not find a significant discontinuity in active investment around the threshold, as our instrument is expected to capture exogenous variation only in passive investment.

I use four measures of price efficiency to examine the impact of passive investment on stock price efficiency in the second stage estimation. First, I construct the pricing error measure of Hasbrouck (1993), which captures the temporary deviation of a transaction price from the (unobservable) efficient price of a security. Second, following Boehmer and Wu (2013), I compute the absolute value of return autocorrelation to capture how closely transaction prices of a security follow a random walk. Third, I construct the lower-frequency price delay measure of Hou and Moskowitz (2005), which captures how quickly prices incorporate market-wide information. Lastly, I use a well-known anomaly in financial markets, the post-earnings announcement drift, as the fourth measure of efficiency (see Ball and Brown (1968)). These unique measures of price efficiency allow us to examine the impact of passive investment on different dimensions of price efficiency: low-frequency, high-frequency, and anomaly.

I find that passive investment, on average, improves the efficiency of security prices. In particular, a one standard deviation increase in passive fund investment in a security is associated with a 0.699, a 0.496, and a 1.103 standard deviation decrease in the pricing error, the absolute value of return autocorrelation, and the price delay measure, respectively. I also find evidence of a decrease in the post-earnings announcement drift as passive fund investment increases. Additionally, I find that stocks with higher active investment exhibit relatively weaker post-earnings announcement drift relative to stocks with lower active investment.

In the main findings, I document a complementary role of passive and active investment on stock price efficiency: the improvement of price efficiency arises for stocks that are analyzed and invested by actively managed funds when information is fully shared with passively managed funds. In particular, I find evidence of a stark improvement in stock price efficiency due to passive investment only when actively managed funds hold significant amounts of the shares outstanding. I sort sample stocks into quartiles based on percent of shares outstanding held by active mutual funds. I find that for stocks in the top quartile, a one standard deviation increase in passive investment is associated with a 0.891, a 0.657, and a 1.214 standard deviation *decrease* in the pricing error, the absolute value of return autocorrelation, and the price delay measure, respectively. However, I do not find evidence of efficiency improvement when active mutual funds have minimal investment For stocks in the bottom quartile, a one standard deviation increase in passive investment is associated with a 0.189 and a 0.215 standard deviation *increase* in the pricing error and the absolute value of return correlation, and a 0.155 standard deviation *decrease* in the price delay measure. All of these changes do not have any statistical significance.

I find evidence that actively managed funds are compensated for their efforts in collecting and processing information. Stocks in the top quartile based on percent of shares outstanding held by actively managed funds deliver significantly higher returns than stocks in the bottom quartile given an equal increase in passive investment. This finding indicates that stocks that are analyzed and invested by actively managed funds compensate those funds with high returns when information is fully shared with passively managed funds.

I also identify economic channels of the efficiency improvement. In particular, I find evidence that stock price efficiency improves due to an increase in analyst following and a decrease in analyst forecast dispersion. Recently, Boone and White (2015) document that higher institutional ownership is associated with greater management disclosure and analyst following, resulting in lower information asymmetry. My empirical analyses reveal that an increase in analyst following and a decrease in analyst forecast dispersion arises for stocks that experience an exogenous passive investment with a significant presence of active mutual funds.

The overall findings in this paper indicate that passive and active mutual funds are complementary to each other in the price discovery process. The impossibility of a perfectly efficient market implies the fact that passively managed funds themselves are not able to make security prices fully efficient because they do not have any incentive to gather and process information. Thus, combined with the significant presence of actively managed funds who exert their effort to gather and process information, security prices become more efficient as information is fully shared with passively managed funds, and actively managed funds are compensated with high returns for their effort.

This paper contributes to several aspects of the literature in finance. First, to the best of my knowledge, this is the only paper which empirically documents the complementarity between passive and active investment in the context of the price discovery process. Academic researchers, as well as practitioners, have focused on the substitutable nature of passive and active mutual funds by studying relative performances of those funds (Fama and French (2010), Elton, Gruber, and Blake (2011)). However, this paper addresses on what ground passive or active mutual funds would prevail in the financial market. On the one hand, if stock prices are significantly inefficient in incorporating information, investors will invest only in actively managed funds until markets become efficient enough. On the other hand, if stock prices are perfectly efficient, no active fund would exist. My findings imply that society requires sizable portions of both passive and active mutual funds for stock prices to be sufficiently efficient.

Second, this paper contributes to the recent literature on the economic consequences of passive investment and composite securities such as ETFs (Exchange Traded Funds). Hamm (2014) finds that ETFs and passive mutual funds deprive the liquidity of the underlying security, and Israeli, Lee, and Sridharan (2017) show that an increase in ETF ownership is associated with reduced price efficiency. Glosten, Nallareddy, and Zou (2016) find evidence that ETF trading actually increase price efficiency for small stocks. My findings suggest that passive investment (or ETF trading) does not play alone in the efficient price discovery process. That is, the efficient price discovery requires both active and passive investment.

Finally, this paper extends the growing literature on the consequences of passive owners in financial markets. When passive owners have the largest stakes of firms, they have strong incentives to influence the governance of a firm (for example, see Appel, Gormley, and Keim (2016)) and information disclosure behavior (Boone and White (2015)). My findings complement this literature by documenting the positive influence of passive investment on the extent to which stock prices reflect information.

The remainder of the paper is organized as follows. Section 2 introduces the various measures of price efficiency used in the main analysis. Section 3 explains background information on the construction of the Russell indexes and my empirical design. Section 4 describes the data. Section 5 provides empirical results. In Section 6, I describe several robustness tests. Section 7 concludes.

2 Measuring Price Efficiency

I employ various approaches to measure how efficiently security prices incorporate information. Two of my main measures of informational efficiency captures how closely transaction prices move relative to random walk using high-frequency transaction data. The third measure that I use in the paper is based on daily returns. This approach considers the speed with which public information is incorporated into prices using daily individual stock and market returns. I also exploit the well-recognized anomaly in financial markets, post-earnings announcement drift (PEAD), to study the impact of passive investment on price efficiency.

2.1 High-frequency efficiency measure

I use two different measures of price efficiency constructed using high-frequency data. The first measure is the pricing error based on Hasbrouck (1993). He assumes that an observed transaction price, p_t, is composed of an unobservable efficient price, m_t, and a pricing error, s_t and that the efficient price is considered as an expected value of a security conditional on all available information. Thus, the pricing error captures the temporary and non-informational related deviation of the transaction price from the efficient price. Following Hasbrouck (1993) and Boehmer and Wu (2013), I estimate a vector autoregressive (VAR) system to separate changes in the efficient price from temporary deviations. In particular, I estimate a dispersion of the pricing error, $\sigma(s)$, from the VAR model as the pricing error is assumed to be a zero-mean and stationary process. In the main analysis, I scale the dispersion with the dispersion of the intraday transaction prices in order for cross-sectional comparison. That is, I refer the ratio of the standard deviation of s to the standard deviation of transaction prices, $\sigma(s)/\sigma(p)$, as the pricing error. Appendix A.1 provides details on the model and estimation.

As a second measure of price efficiency, I use the absolute value of intraday return autocorrelation. Intuitively, if security prices are perfectly efficient in incorporating information, the movement of prices should follow a random walk. This measure is computed from intraday transaction data and captures temporary deviation from a random walk. Thus, if the investment of passive mutual funds improves the price efficiency, the transaction prices should exhibit low autocorrelation in either direction resulting in the small absolute value of autocorrelation. Similar to Boehmer and Wu (2013) and Chordia, Roll, and Subrahmanyam (2005), I choose a thirty-minute interval to estimate return autocorrelation of transaction prices and denote |AR30| as the absolute value of the autocorrelation.

2.2 Low-frequency efficiency measure

For a low-frequency measure of price efficiency, I construct a price delay measure introduced by Hou and Moskowitz (2005), which captures the speed of adjustment of an individual stock to incorporate market-wide information. If today's stock prices cannot fully incorporate information due to inefficiency, remaining information will be gradually absorbed into prices. Based on this intuition, the price delay is estimated from a market model regression that is extended to include the lagged market returns. Griffin, Kelly, and Nardari (2010) and Saffi and Sigurdsson (2011) apply the price delay measure in an international context.

The original price delay measure suggested by Hou and Moskowitz (2005) is an annual frequency using weekly returns in the estimation. However, it is likely that the impact of passive funds investment is concentrated around the time of reconstitution of the Russell indexes. Thus, using an annual frequency measure might not be precise enough to capture the impact of passive funds on price efficiency. Following Boehmer and Wu (2013), I modify the approach of Hou and Moskowitz (2005) and compute monthly price delays using daily returns, contemporaneous market returns, and five days of lagged market returns as the following regression:

$$r_{i,t} = \alpha_i + \beta_i r_{m,t} + \sum_{n=1}^{5} \delta_i^{-n} r_{m,t-n} + \varepsilon_{i,t}, \tag{1}$$

where $r_{i,t}$ is the return from stock i on day t and $r_{m,t}$ is the market return on day t. Then, I estimate a second regression that restricts the coefficients on lagged market returns to zero. The price delay measure is calculated as one minus a ratio of the R-squared from the restricted model over the R-squared from the unrestricted model:

$$Price\ Delay = 1 - \frac{R^2_{\delta_i^{-n}=0, \forall n \in [1,4]}}{R^2}. \tag{2}$$

That is, the price delay measure captures the fraction of variation of contemporaneous individual stock returns explained by lagged market returns.

2.3 Post-earnings announcement drift

Since Ball and Brown (1968) documented that abnormal returns of stocks with positive (negative) earnings surprises tend to exhibit positive (negative) subsequent returns for several weeks following the earnings announcement, this well-established phenomenon, post-earnings announcement drift (PEAD), indicates some degree of informational inefficiency in the financial markets. If an exogenous increase in passive investment on a stock impairs its price efficiency due to a lack of security analysis or monitoring, the stock which experienced an increase in passive investment should exhibit strong post-earnings announcement drift. However, if an increase in passive investment improves price efficiency through corporate governance or quality revelation, the post-earnings announcement drift should be attenuated.

I compute earnings surprises as the difference between actual earnings and the most recent I/B/E/S consensus forecasts of analysts. Then I construct post-earnings announcement drift as cumulative abnormal returns over five- and ten-day windows starting from the second trading day after the earnings announcement. Abnormal returns are computed as a stock's raw returns net of value-weighted CRSP returns.

3 Empirical Design

The empirical approach of this paper consists of two stages. In the first step, I separately investigate the impact of passive investment on stock price efficiency using an instrument to capture an exogenous increase in passive investment. In the second step, changes of price efficiency depending on the mix of passive and active investment are examined. Actively managed funds try to outperform a market or benchmark index by researching and investing in individual stocks with inefficient prices. Whereas passively managed funds mechanically track their benchmark index because they believe stock prices are efficient so that it is impossible to outperform the market return. However, a recent dramatic growth and predominance of passive mutual funds casts doubt on their contribution to stock price efficiency, because their main objective is to match the performance of a market index by holding a basket of representative stocks in the index in proportion to their weights in the index.

Examining how passive investment affects stock price efficiency, is empirically challenging due to several reasons. First and foremost, but often ignored in the asset pricing literature, a causal relationship between passive investment and stock price efficiency is subject to a severe endogeneity issue. That is, investment by passive mutual funds could be correlated with other factors, such as transaction cost or information asymmetry. For example, Falkenstein (1996) documents that mutual funds have a strong preference to stocks with low transaction cost, high liquidity, and low idiosyncratic volatility. Second, an ordinary least squares estimation of passive investment on stock price efficiency is subject to omitted variable bias. It is still questionable whether observable control variables and fixed effects can fully capture characteristics that simultaneously determine passive investment and stock price efficiency. To overcome this empirical challenge, I use the index assignment of firms into the top of the Russell 2000 (annual reconstitution) as exogenous variation in passive mutual fund investment.

3.1 Russell Indexes Construction

The Russell 1000 index consists of the largest 1,000 U.S. listed firms, while the Russell 2000 index comprises the subsequent 2,000 largest firms. These indexes provided and maintained by FTSE Russell represent approximately 98% of the entire public equity market in the U.S, and are widely used as benchmarks. Chang, Hong, and Liskovich (2015) document that the dollar amount of institutional assets benchmarked to the Russell 1000 index is $90 billion, while the Russell 2000 index is tracked by around $200 billion.

Reconstitution of these Russell indexes provides us a clean empirical laboratory to examine the impact of passive investment on price efficiency by generating an exogenous shock to passive mutual fund holdings. As the first step of the reconstitution, every year on the last trading day of May, stocks are ranked by their market capitalizations.[6] Second, on the last Friday of June, the indexes are reconstituted such that firms ranked from 1st to 1,000th and firms ranked from 1,001st to 3,000 constitute the Russell 1000 and Russell 2000 indexes, respectively. Each stock's weight in the index is determined by its float-adjusted market capitalization at the end of June. The float adjustment accounts for the value of shares that are not publicly available.[7]

Since 2007, Russell has adopted a banding policy to mitigate turnover of members in the indexes and to reduce unnecessary trading. Under the banding policy, firms with a certain range of the cutoff would not switch indexes unless the market capitalization of a firm deviates far enough to ensure an index membership change. In particular, all stocks included in the Russell 1000 and 2000 in the previous year are ranked from smallest to largest at the end of May, then a cumulative market capitalization is computed for every stock. This cumulative market capitalization of each stock is expressed as a percentage of the total market capitalization of all stocks in Russell 1000 and 2000 indexes. Based on these values, stocks switch from their current index only if they move beyond a 5% band around the cumulative market

[6] ADR, ADS, preferred stocks, redeemable shares, warrants, rights, and trust receipts are excluded.

[7] For example, shares held by a government or by an employee stock ownership plan will be excluded when the Russell calculates a firm's float-adjusted market capitalization. Detailed mechanism of the float-adjustment made by the Russell is unknown. This is why I control for the float-adjustment using other proxy, which will be explained in Section 3.2.

capitalization of the 1,000th stock. I use market capitalization obtained from the CRSP to compute these values and the implied cutoffs for each year from 2007 to 2016.[8]

The index assignment has a significant effect on portfolio weights, in particular for stocks assigned to the top of the Russell 2000 index, because the Russell indexes are value-weighted. For example, the 1,001st largest stock at the end of May in 2006 will be assigned to the Russell 2000 index and be given a very large weight in the index once the annual reconstitution is completed at the end of June, while the 1,000th largest stock will be assigned to the Russell 1000 index and be given a very small weight in the index. Figure 1 plots weights of stocks in their indexes around the cutoff in 2006 and shows a significant difference of portfolio weights between stocks in the bottom 250 of the Russell 1000 index and those in the top 250 of the Russell 2000 index. While the average portfolio weight of stocks in the top 250 of the Russell 2000 index is around 0.123%, the average weight of the bottom 250 stocks of the Russell 1000 index is around 0.015%. That is, the stocks assigned in the top 250 of the Russell 2000 index are given almost 8.2 times greater weights in the index than those assigned in the bottom 250 of the Russell 1000 index.

The index assignment, which causes the difference in portfolio weights of stocks in the indexes, further impacts the investment of passive mutual funds. Passive mutual funds aim to minimize tracking error in mimicking their benchmarked indexes by adjusting their holdings based on weights of stocks in the indexes. Thus, it is important for passive mutual funds to match their holdings according to weights for stocks in the top of the index because those stocks are more likely to influence the overall performance of the index than stocks in the bottom of the index when the benchmarked index is value-weighted. Even some mutual funds tracking an index could choose to hold a few representative stocks in the index based on their weights and exclude some stocks in the bottom of the index (see Frino and Gallagher (2001)). For example, if a stock is assigned into the top 250 of the Russell 2000 index from the bottom

[8] Chang, Hong, and Liskovich (2015) also computed these implied cutoffs each year to examine the price effects of index inclusion, while Appel, Gormley, and Keim (2016) and Crane, Michenaud, and Weston (2016) drop observations after 2006. Main analyses presented in the paper use the implied cutoffs, but all results are robust if I drop observations from 2007.

250 of the Russell 1000 index, passive mutual funds tracking the Russell 2000 index would significantly increase their holdings of the stock in order to minimize tracking error.

Figure 2 visually highlights the impact of index assignment on passive mutual fund investment. In Figure 2, I plot the average percentage of shares outstanding held by all (top panel), passive (middle panel), and active mutual funds (bottom panel) to total shares outstanding of firms over 100 bins across all years. The x- and y-axes represent a firm rank of weight in the index at the end of June and an average percentage holdings by each type of mutual fund at the end of September, respectively. The figure displays a large discontinuity in the percentage of passive mutual fund holdings (middle panel), while the average percentage holdings by active mutual funds do not exhibit any discontinuity around the cutoff. As Chang, Hong, and Liskovich (2015) show that there is no structural break with respect to firm characteristics (size, ROE, ROA, EPS, etc.) around the index cutoff, the discontinuity in the passive mutual fund holdings is due to the index assignment causing differences in weights between the top of the Russell 2000 index and the bottom of the Russell 1000 index. This allows us to use the annual reconstitution of the Russell indexes as a valid instrument to capture exogenous variation in passive mutual fund investment.

3.2 Empirical Specification

To formally test the impact of passive investment on stock price efficiency in the first step, I use an identification strategy using inclusion in the Russell 2000 index as an instrument for passive fund investment. Once I identify the impact of an exogenous increase of passive investment on price efficiency, I further investigate the role of active investment in conjunction with an increase of passive investment in the second step. In particular, I examine price efficiency measures in a narrow bandwidth around the cutoff as a function of instrumented passive fund holdings following Lee and Lemieux (2010). In the first stage of the estimation, I capture exogenous variation in passive investment using the instrument, the Russell index assignment. That is, the first stage regression is as below:

$$\begin{aligned}Passive\%_{i,t} = &\tau Russell2000_{i,t} + \delta_1(Rank^*_{i,t} - c) + \delta_2 Russell2000_{i,t}(Rank^*_{i,t} - c)\\ &+ \delta_3 FloatAdj_{i,t} + \delta_4 Liquidity\ Controls + \alpha_t + \theta_i + \varepsilon_{i,t},\end{aligned} \tag{3}$$

where $Passive\%_{i,t}$ is the percentage of shares held by passive mutual funds. For mutual fund holdings, I use the reports of funds filed in S12 mutual fund holdings database at the end of September in year t, which is the first quarter-end after annual reconstitution of the Russell indexes. $Russell2000_{i,t}$ is a dummy variable equal to one if a firm i is included in the Russell 2000 index in year t as of the end of June. $Rank^*_{i,t}$ is the rank of a firm based on market capitalization at the time of index assignment.[9] c is a cutoff of the Russell 1000 index, which is 1000 before banding policy and is calculated separately every year after banding policy is implemented. Russell uses a proprietary float-adjustment process that results in firms with low floating shares ranked lower in an index than predicted by their market capitalization as of the end of May. Thus, I construct a variable, $FloatAdj_{i,t}$, as a difference between the rank implied by the end-of-May capitalization and the actual rank assigned in the index by the Russell at the end-of-June of firm i in year t and is used as a proxy for the adjustment made by Russell for floating shares. I also control for liquidity effect by adding proxies for liquidity (*Liquidity Controls* include Amihud (2002) measure and zeros introduced by Goyenko, Holden, and Trzcinka (2009). The construction of these variables are explained in detail in Appendix A.2)[10], as Boone and White (2015) document that higher institutional ownership is associated with higher trading volume and liquidity.

In the second stage regression, I estimate the impact of instrumented passive fund hol-

[9] I do not use actual rankings or weights of stocks because of potential endogeneity concerns about unobserved determinants of actual weights assigned by Russell. Chang, Hong, and Liskovich (2015) discuss about reasons why the actual weights or rankings should not be used as an instrument.

[10] Goyenko, Holden, and Trzcinka (2009) examine and run horseraces of various measures of liquidity. They find that the illiquidity measure of Amihud (2002) and zeros (the proportion of positive-volume days with zero returns) outperform in capturing the price impact.

dings on various measures of price efficiency.

$$\begin{aligned}Efficiency_{i,t} = \beta \widehat{Passive\%}_{i,t} + \gamma_1(Rank^*_{i,t} - c) + \gamma_2 Russell2000_{i,t}(Rank^*_{i,t} - c) \\ + \gamma_3 FloatAdj_{i,t} + \gamma_4 Liquidity\ Controls + \kappa_t + \eta_i + \epsilon_{i,t},\end{aligned} \tag{4}$$

where $Efficiency_{i,t}$ are various price efficiency measures for firm i in year t. As the mutual fund holdings are measured at the end of September, I take an average of a variable from July (the first month after the reconstitution) to September in year t.

The key feature of the empirical design is to identify exogenous variation in passive investment, which I examine around the Russell 2000 inclusion threshold. To identify variation around the threshold, I control for the distance to the threshold of observed market capitalization ranking, $(Rank^*_{i,t} - c)$ as well as for the interaction $Russell2000_{i,t}(Rank^*_{i,t} - c)$ of firm i in year t.[11] Thus, my key instrument is $Russell2000_{i,t}$, conditional on market capitalization ranking, $(Rank^*_{i,t} - c)$, and the interaction $Russell2000_{i,t}(Rank^*_{i,t} - c)$. Both regressions include year and firm fixed effects. All standard errors from the estimation of the above regressions are clustered at the firm level.

3.3 Optimal Bandwidth

I use regression discontinuity around the Russell indexes cutoff with an instrument estimation to examine the impact of passive investment on price efficiency. Thus the choice of bandwidth, i.e., how many firms on either side of the cutoffs are used in the estimation, is another variable to be determined. The choice should balance the benefits of more precise estimates as the sample size grows and the costs of increased bias. Chang, Hong, and Liskovich (2015) use a bandwidth of 100 around the cutoff in estimating the deletion and addition effect of indexing. Appel, Gormley, and Keim (2016) use a bandwidth of 250 in their main analysis on the influence of passive owners, while Crane, Michenaud, and Weston (2016) provide the results

[11] $(Rank^*_{i,t} - c)$ and $Russell2000_{i,t}(Rank^*_{i,t} - c)$ control for the mechanical relationship with market capitalization ranking on either side of the cutoff. Thus, they isolate any difference in passive mutual fund holdings around index inclusion at the cutoff.

for the bandwidth of 100 and 500 in examining the effect of institutional ownership on payout policy.

To determine the optimal bandwidth for the main analysis, I use the optimal rule of thumb bandwidth selection procedure prescribed in Calonico, Cattaneo, and Titiunik (2014). Over the full sample period, the optimal bandwidth estimated using this process is 276. While the optimal bandwidth over the period from 1996 to 2006 (before the banding policy was implemented) is 118, the bandwidth from 2007 to 2016 (after the banding policy was implemented) is 341. In my main analysis, I use a bandwidth of 250 around the cutoff. I test and confirm the robustness of findings using the bandwidths of 100 and 500 reported in Section 6.

4 Data

4.1 High-, low-frequency, earnings announcement, and analyst data

Two measures of price efficiency explained in Section 2.1 are constructed using high-frequency transaction data. I collect high-frequency data on security prices from the Trade and Quote (TAQ) database.[12] One of price efficiency measure, which captures how fast market-wide information is incorporated into security price, uses daily returns. Thus, I obtain daily stock prices and CRSP value-weighted market returns from the CRSP database. When I construct earnings announcement surprises and post-earnings announcement drift, I use both I/B/E/S and COMPUSTAT databases. Data to construct analyst following and analyst forecast dispersion is obtained from I/B/E/S. Other accounting variables are obtained from the COMPUSTAT database.

[12] Due to the limitation of data subscription, high-frequency data is only available from January 2001 to December 2014.

4.2 Mutual fund holdings data

My sample consists of the Russell 1000 and 2000 indexes constituents from 1996 to 2016. I obtain mutual fund ownership data for the sample firms from the S12 mutual fund holdings data provided by Thomson Reuters. All mutual funds in the U.S. are required to report their stock holdings to the Securities and Exchange Commission. Before 2004, funds were required to report holdings twice a year, but many mutual fund voluntarily reported their holdings other two quarters. However, since 2004, all mutual funds must report the holdings every quarter. There are multiple mutual funds that report their holdings more than once in a given month. For those funds, I keep only the last report of the month. As I use the Russell Indexes reconstitution as an instrument to capture exogenous variation of passive investment, I collect mutual fund holdings reports from the S12 database at the end of September, which is the end of the first quarter after annual reconstitution (at the end of June) of the Russell indexes.

To classify a mutual fund as passively or actively managed, I follow a method used by Iliev and Lowry (2015) and Appel, Gormley, and Keim (2016). Using the linking file provided by Wharton Research Data Services (WRDS), I merge the mutual fund holdings data from Thomson Reuters with the Center for Research in Security Prices (CRSP) mutual fund data, which contains detailed information on fund names, investment objectives, management companies, and so on. From the merged data, I identify a passively managed fund if a name of the fund contains a string that represents it as an index fund, or if an investment objective code in the database classifies the fund as an index or a passively managed fund. All other funds are classified as actively managed funds using their investment objective codes in the CRSP mutual fund database, and the remaining funds that cannot be classified are left unclassified. For these three types of mutual funds holdings data, I further collect each firm's market capitalization data from the CRSP database by multiplying the number of shares outstanding with the monthly closing price of a security. I then compute the percentage of a firm's market capitalization owned by passively, actively, and unclassified funds.

4.3 Sample and descriptive statistics

Table 1 provides descriptive statistics of key variables for the main analysis. Panel A and Panel B report statistics for all firms included in the Russell 1000 and Russell 2000 indexes and firms in the 250 bandwidth around the cutoff of the indexes, respectively. For any firms once included in the Russell 1000 or 2000 indexes in the sample period from 1996 to 2016 (Panel A), the average percentage of shares outstanding held by mutual funds is around 15%, and active mutual funds represent the largest portion of the mutual fund holdings at around 9%. Passive mutual funds and unclassified mutual funds account for 2.6% and 3.4% of shares outstanding, respectively. The average size of firms in my sample is around $4.9 billion.

Panel B of Table 1 reports descriptive statistics for a restricted sample of firms in a 250 bandwidth for the main analysis. Total mutual fund holdings of stocks around the bandwidth are slightly higher compared to those in Panel A. The average percentage of all mutual fund holdings is around 17.6%, and passive, active, and unclassified mutual fund holdings account for 2.9%, 10.7%, and 4.0%, respectively. As firms in Panel B represent mid- and small-cap stocks, an average size of firms, $1.7 billion, is smaller than that of firms in Panel A, which consist of all stocks in the Russell 1000 and Russell 2000 indexes. Descriptive statistics for the main measures of price efficiency, the absolute value of return autocorrelation and the price delay measure, are comparable to Panel A, suggesting that the size of a firm is not a decisive factor of stock price efficiency.

5 Empirical Results

5.1 Passive Investment and Price Efficiency

In this section, I provide one of the main results of the paper about the impact of passive investment on price efficiency using the Russell 2000 index inclusion as an instrument in the regression discontinuity design. While the main analyses use a bandwidth of 250 as described in Section 3.3, in Table 2 I report descriptive statistics of key variables for firms in the bottom of the Russell 1000 index and the top of the Russell 2000 index for different choices of bandwidths. For ±100, ±250, and ±500 firms around the threshold of the Russell indexes, Panel A, Panel B, and Panel C, respectively, report an average, median, and standard deviation of variables.

As expected for a valid instrument to capture exogenous variation in passive fund investment, I observe higher mutual fund holdings on stocks in the top of the Russell 2000 than those in the bottom of the Russell 1000, which is largely due to higher holdings from passive mutual funds. For example, in Panel B of Table 2, the top 250 firms in the Russell 2000 index have 2.32% greater aggregate mutual fund holdings compared to the bottom 250 firms in the Russell 1000 index; almost half of this difference results from greater holdings by passive funds. Passive investment is about 33% higher for firms in the top 250 of the Russell 2000 index relative to firms in the bottom 250 of the Russell 1000 index. From the descriptive statistics, I find less pricing error and price delay for stocks in the top of the Russell 2000 index. I also observe higher liquidity (lower *Amihud* and *zeros*) for stocks in the top of the Russell 2000, and this pattern confirms the finding of Boone and White (2015) that an increase of institutional ownership improves trading volume and liquidity. Thus, I explicitly control for liquidity in the main analyses. In the following subsections, I formally test and document the impact of passive investment using my identification strategy.

5.1.1 First-stage estimation: Passive investment around the index threshold

In this subsection, I report estimates of the first-stage regression of passive mutual fund holdings on the Russell 2000 inclusion around the threshold, conditional on market capitalization ranking and the interaction between the inclusion and the ranking. In particular, I estimate the following equation:

$$\begin{aligned} \mathit{MF\ Holdings\%}_{i,t} = {} & \tau \mathit{Russell2000}_{i,t} + \delta_1(\mathit{Rank}^*_{i,t} - c) + \delta_2 \mathit{Russell2000}_{i,t}(\mathit{Rank}^*_{i,t} - c) \\ & + \delta_3 \mathit{FloatAdj}_{i,t} + \delta_4 \mathit{Liquidity\ Controls} \alpha_t + \theta_i + \varepsilon_{i,t}, \end{aligned} \tag{5}$$

where $MF\ Holdings\%_{i,t}$ is the percentage of firm i's shares at the end of the first quarter (end of September) of year t held by different categories of mutual funds: all mutual funds, passively managed funds, actively managed funds, and unclassified mutual funds. Other variables are explained in detail in Section 3.2.

The estimation results are provided in Table 3, confirming that the inclusion of a stock in the Russell 2000 index is strongly associated with an increase in passive fund investment. The statistical significance remains strong even after controlling for liquidity by adding the illiquidity measure of Amihud (2002) and the proportion of days with positive-volume and zero returns (*zeros*). While the graphical analysis in Figure 2 shows stark differences in mutual fund holding around the index threshold, especially in passively managed fund holdings, the results in Table 4 provide point estimates of the causal effect of the index inclusion on mutual fund investment. Column (2) in Table 4 shows that passive mutual fund holdings are significantly higher for stocks in the top 250 of the Russell 2000 index than for those in the bottom 250 of the Russell 1000 index. In particular, those firms just included in the top of the Russell 2000 index have 33.5 percentage point more shares held by passively managed funds, and this difference is statistically significant at the 1% level. However, in column (1) of Table 4, aggregate mutual fund holding does not show a statistically significant difference between stocks in the top of the Russell 2000 and those in the bottom of the Russell 1000.[13] As other types of mutual funds, including actively managed funds, do not have a strong incentive or motivation to mechanically track an index portfolio, I do not observe any discontinuity in mutual fund holdings by active mutual funds and unclassified funds around the Russell index threshold.

[13] Appel, Gormley, and Keim (2016) show that aggregate mutual fund ownership is significantly (at the 10% level) higher for 250 stocks in the top of the Russell 2000. A reason why I do not find a significant difference is that I (1) use a different regression specification similar to Crane, Michenaud, and Weston (2016) and (2) include both year and firm fixed effects. Thus, the estimates identify within-year and -firm variation of passive investment depending on the Russell 2000 index inclusion.

5.1.2 Impact of passive investment on price efficiency

In this subsection, I examine how passive fund investment affects the efficiency of security prices. From this point, I scale both price efficiency measures ($\mathit{Efficiency}_{i,t}$) and passive fund holdings percentage ($\mathit{Passive}\%_{i,t}$) by their sample standard deviations so that the point estimate of β in Equation (4) can be interpreted as the standard deviation difference in a price efficiency measure for one standard deviation increase in $\mathit{Passive}\%_{i,t}$.

Table 4 reports the two-stage least-squares estimates of passive fund holdings on price efficiency measures described in Equations (3) and (4). Panel A and Panel B provide the results for the first-stage (Equation(3)) and second-stage (Equation(4)), respectively. The first-stage estimates using scaled variables by their sample standard deviation confirm that stocks assigned into the top 250 of the Russell 2000 index have significantly (at the 1% level) higher ownerships by passively managed funds.[14] The point estimate in column (1), for example, shows an increase in passive mutual fund holdings of about a half of a sample standard deviation.

The results of the second-stage regression are provided in Panel B of Table 4. I find that investment of passive mutual funds has a positive impact on my measures of price efficiency. The coefficient estimates on the instrumented passive mutual fund holdings in Equation (4) are statistically negative at the 1% level for all three price efficiency measures: the pricing error of Hasbrouck (1993) (column (1)), the absolute value of autocorrelation (column (2)), and the price delay of Hou and Moskowitz (2005) (column (3)). In particular, one standard deviation increase in passive fund investment on a security is associated with a 0.699, a 0.496, and a 1.103 standard deviations decrease in the pricing error, the absolute value of return autocorrelation, and the price delay, respectively. That is, the finding suggests that an exogenous increase of passive investment improves the efficiency of security prices. This finding is robust to different choices of bandwidths and specification, which is also provided in Section 6.

The main results in Table 4 are robust to the implementation of a banding policy of

[14]For all columns in Table 4, F-statistics and t-statistics exceed the thresholds suggested by Stock and Yogo (2005).

Russell. As Russell implements the policy to reduce turnover of stocks inclusion and deletion in the Russell 1000 and 2000 indexes from 2007, I first confirm the robustness of the first-stage estimation results before and after the banding policy implementation. In an unreported analysis, the first-stage regressions are quantitatively similar before and after banding, which indicates that the estimation of the implied threshold from 2007 is accurate. Also, due to the robustness of the first-stage results, I also find that second stage results are similar before and after banding.

5.2 Passive, Active Investment, and Price Efficiency

In this subsection, I consider a role of active investment in the association between passive investment and price efficiency improvement. Grossman and Stiglitz (1980) argue that perfectly informationally efficient markets are impossible because information is costly. If a market is perfectly efficient, the compensation to information gathering and processing is zero. Alternatively, the degree of inefficiency determines the effort that investors are willing to gather and trade on the information. In the mutual fund industry, actively managed funds charge relatively higher fees and expenses than passively managed funds as a compensation for their effort on security analysis. Given the impossibility of a perfectly efficient market and a presence of actively managed funds, there exists opportunities to obtain excess returns until information is fully reflected in stock prices. That is, the presence of actively managed funds would be a key determinant of the extent to which stock prices reflect information.

I investigate any significant differences in price efficiency measures in regard to a mix of passive and active mutual fund investment. I expect to observe high price efficiency for stocks with high passive investment as well as high active investment. To examine this, I conduct a double-sorting analysis. That is, I first sort all sample firms into tercile portfolios based on the percentage of shares held by passive mutual funds. Then, for each tercile, firms are sorted into tercile portfolios based on the percentage of shares held by active mutual funds. I examine three different price efficiency measures for double-sorted portfolios.

I find evidence that the three price efficiency measures (pricing error, absolute value of autocorrelation, and price delay) are significantly lower when both passive and active funds share a significant portion of the company's stock. Table 5 reports averages of price efficiency measures for the bottom (*Low*) and top (*High*) tercile portfolios and the differences of averages between the top and the bottom terciles (*Diff(H-L)*) along with their t-statistics. In Panel A of Table 5, the average pricing error for firms in the bottom terciles of both active and passive investment is 0.1692, while the average for firms in the top terciles of both active and passive investment is 0.1029. The differences in all price efficiency measures are statistically significant. Thus, I observe that security prices are more efficient for stocks whose shares are largely held by both passive and active funds.

In the regression framework, I also find evidence that the presence of actively managed funds plays a critical role in the price efficiency improvement from passively managed funds. In particular, I find that an exogenous increase in passive investment causes an improvement in price efficiency only when there exists enough shares held by actively managed funds. To examine the role of active investment in the regression framework, I sort sample firms into quartiles based on their percentage holdings owned by actively managed funds each year and examine the impact of passive investment on price efficiency measures for each quartile using Equations (3) and (4).

Table 6 reports the results for the second-stage estimation.[15] Panel A, Panel B, and Panel C show the estimation results for the pricing error of Hasbrouck (1993), the absolute value of autocorrelation, and the price delay of Hou and Moskowitz (2005), respectively. The estimated coefficients on the instrumented passive investment are positive and statistically significant only for the top quartiles of active fund holdings at the 1% level. The estimates indicate that, for example, when a stock is largely held by actively managed funds (top quartile), a one standard deviation increase in *Passive%* is associated with a 0.891 standard deviation decrease in the pricing error, a 0.657 standard deviation decrease in the absolute

[15]The first-stage regressions results for all quartiles are similar to the results in Table 4, and F-statistics and t-statistics exceed the suggested threshold by Stock and Yogo (2005).

value of autocorrelation, and a 1.214 standard deviation decrease in the price delay. However, if a stock is rarely held by active mutual funds (bottom quartile), passive investment does not improve any stock price efficiency.

I next analyze whether those stocks largely held by actively managed funds compensate active funds for their effort on information gathering and processing. To test whether active funds are compensated, I sort all sample firms into quartiles based on percentage shares held by actively managed funds each year, and examine cumulative returns and cumulative trading volume from July to September of a corresponding year.[16] In particular, I estimate a similar two-stage IV regression with replacing a dependent variable in Equation (4) to cumulative returns or cumulative trading volumes.

Table 7 provides the second-stage regression results of cumulative returns (Panel A) and cumulative trading volumes (Panel B). I find evidence that stocks whose shares are largely held by actively managed funds deliver higher cumulative returns and lower cumulative trading volumes once stocks are included in the top 250 of the Russell 2000 indexes, relative to stocks with minimal active investment. In Panel A, the coefficient estimates on *Passive%* are positive and statistically significant (at the 10% level) in columns (3) and (4), indicating higher cumulative returns for stocks with high active fund holdings. In Panel B, the coefficient estimates on *Passive%* monotonically decrease with the percentage of active fund holdings, and a difference of estimates between the top and the bottom quartiles is statistically significant. The findings in Table 7 suggest that active mutual funds maintain their holdings and are compensated with high returns from stocks experiencing an increase of passive investment, as information on stocks included in the index with significant weights is fully revealed by passive funds and other market participants.

[16] I do not use abnormal returns or trading volumes, because there is no systematic difference in firms around the index threshold. See, for example, Chang, Hong, and Liskovich (2015) and Crane, Michenaud, and Weston (2016).

5.3 Passive/Active investment and post-earnings announcement drift

In this subsection, I analyze how passive investment is associated with post-earnings announcement drift. Ball and Brown (1968) first document that returns tend to be positive after positive earnings surprises and negative after negative surprises, indicating that stock prices do not fully and immediately incorporate information at the time of the announcement. Bernard and Thomas (1989) find that post-earnings announcement drift is a manifestation of investors' failure to recognize the information in the earnings surprises. Thus, if the investment of passive and active funds affects price efficiency as described in previous sections, the investment would affect post-earnings announcement drift in similar manners. To examine the impact of passive and active investment on post-earnings announcement drift, I examine two weeks following the earnings announcement. I focus earnings announcements of firms between July and September and sort them into quartiles based on earnings surprises. Then, I investigate returns on the first trading day after the announcements and cumulative returns over one- and two-week windows.

I first find that post-earnings announcement drift prevails in not only all firms included in either the Russell 1000 or 2000 indexes but also in firms around the index cutoff. Table 8 reports, for each quartile of earnings surprises, average abnormal returns on the first trading days (column under *Announcement Day*), average cumulative returns from one-day after the first trading day over 5 trading days (columns under $[+1, +5]$) and over 10 trading days (column under $[+1, +10]$) for all firms in the Russell 1000 and 2000 indexes (columns under *Full Sample*) and for firms in a 250 bandwidth around the index threshold (columns under *Bandwidth=250*). For example, for firms in the 250 bandwidth with the most negative earnings surprises (quartile 1), an average abnormal return on the first trading day after the announcements is -0.426% (with t-statistics of -27.989). Firms in the bandwidth with the most positive earnings surprises (quartile 4) exhibit significantly positive abnormal returns on the first trading day of the announcements with 0.492% (with t-statistics of -36.904). This finding on significant abnormal returns on the announcement confirms that earning announcement effects are very strong in my sample. I also observe significant post-earnings announcement drift, which is the strongest for the most negative and the most positive earnings surprise portfolios. While 5-trading day cumulative abnormal returns starting from the second trading day after the announcement date $[+1, +5]$ are -0.745% (with t-statistics of -6.483) for stocks with the most negative earnings surprises and 6.47% (with t-statistics of 5.859) for stocks with the most positive surprises.

I now examine how post-earnings announcement drift is affected by passive and active investment. In this analysis, I only focus on two quartiles with extreme earnings surprises (quartile 1 and 4) as they exhibit the most significant post-earnings announcement drift. I find evidence that, in general, a degree of post-earnings announcement drift decreases as stocks have a greater amount of passive mutual fund investment. In addition, I find that stocks with higher active investment exhibit relatively smaller earnings surprises. Table 9 reports the analysis on post-earnings announcement drift. In Panel A, I split firms around the index cutoff (with a bandwidth of 250) based on their assignment in the Russell indexes. That is, in column (1) and column (2), I report 5-trading day cumulative abnormal returns for the most positive and the most negative earnings surprises quartiles for firms included in the bottom 250 of the Russell 1000 index and firms included in the top 250 of the Russell 2000 index. In the last column, differences between column (1) and column (2) are reported with t-statistics. I find that a degree of post-earnings announcement drift is smaller for stocks included in the top of the Russell 2000 index than for stocks in the bottom of the Russell 1000 index, and the difference is statistically significant at the 1% level for both the most negative and the most positive earnings surprise quartiles.

In Panel B and Panel C, I further split firms in Panel A into stocks in the top quartile of active fund holdings and stocks in the bottom quartile of active fund holdings, respectively. I find evidence that for both samples of firms, post-earnings announcement drift is less pronounced for firms in the top of the Russell 2000 than for those in the bottom of the Russell 1000 index. However, it is noteworthy that a degree of post-earnings announcement drift is much weaker when stocks are largely held by actively managed funds. Comparing Panel B with Panel C in Table 9, the magnitude of post-earnings announcement drift, both in column (1) and column (2), is much smaller for stocks with high active holdings, even though differences are both significant in the last column. This finding suggests that investment of active mutual funds plays a complementary role in price efficiency improvement along with passive investment, consistent with my main finding in Section 5.1.

5.4 Economic Channels of Price Efficiency Improvement

In this subsection, I examine possible economic channels of the efficiency improvement. Analyst play an important role as information intermediaries by gathering and processing information about firms. Recently, Boone and White (2015) document that higher institutional ownership is associated with greater analyst following and lower analyst forecast dispersion, resulting in lower information asymmetry. Thus, I expect that analysts contributes the efficiency improvement for stocks experiencing both passive and active mutual fund investment, documented in the previous subsections.

Using the regression discontinuity framework used in Section 5.2, I investigate the effect on analyst following and analyst forecast dispersion of passive fund investment depending on the investment of active mutual funds.[17] I find that an exogenous increase in passive investment causes an increase in analyst following and a decrease in analyst forecast dispersion only when there exists significant shares held by active mutual funds. Table 10 reports the second-stage regression results of analyst following (Panel A) and analyst forecast dispersion (Panel B) on passive fund investment for each quartile based on active fund investment. In particular, for stocks in the top quartile of active mutual fund investment, a one standard deviation increase of passive fund investment is associated with 1.336 more unique analysts providing one-year-ahead annual forecast and a 0.387 standard deviation decrease in analyst forecast dispersion. However, I do not find any statistical association of analyst following and

[17] *Analyst Following* is constructed as the number of unique analysts providing one-year-ahead annual forecasts, and *Analyst Forecast Dispersion* is constructed as the standard deviation of the consensus one-year-ahead annual earnings estimates divided by the absolute value of the mean consensus estimate. These monthly variables are averaged from July to September every year.

analyst forecast dispersion with passive fund investment for other lower quartiles.

6 Robustness Check

6.1 Alternative Specification

There exists some confusion in the literature about how best to exploit the Russell index setting and whether to use an instrumental variable estimation or fuzzy or sharp regression discontinuity design. Theoretically, a discontinuity design is called the sharp regression design if a treatment is known to depend in a deterministic way on some observable variables, while one is called the fuzzy regression design if a treatment is a random variable given as a conditional probability of some observable variables (see, for example, Lee and Lemieux (2010) and Hahn, Todd, and Van der Klaauw (2001)). My approach exploits a sharp regression discontinuity design using an instrument as the main variables of interest are mutual fund holdings and various measures of price efficiency after the Russell reconstitution. That is, a sharp regression discontinuity setting is appropriate for my analysis because I am able to observe actual assignments of firms into the indexes at the time of analysis. Chang, Hong, and Liskovich (2015) use a fuzzy regression discontinuity design in investigating returns due to buying and selling between the ranking date (the end of May) and the reconstitution date (the end of June). Their choice of a design is appropriate because a treatment (the index inclusion) can only be predicted using a market capitalization at the end of May.

I consider other empirical designs and confirm that my finding is robust to different specifications. Appel, Gormley, and Keim (2016) examine how passive owners affect the governance of a firm using the Russell index setting. Thus, for another possible specification, I examine my main finding using the regression specification of Appel, Gormley, and Keim

(2016). In particular, I estimate the following first-stage regression:

$$\begin{aligned} \mathit{Passive\%}_{i,t} = \tau \mathit{Russell2000}_{i,t} + \sum_{n=1}^{N} \delta_n (\log(\mathit{Market\ Cap}_{i,t}))^n \\ + \delta_{N+1} \mathit{FloatAdj}_{i,t} + \alpha_t + \theta_i + \varepsilon_{i,t}, \end{aligned} \tag{6}$$

where $\mathit{Market\ Cap}_{i,t}$ is the end-of-May CRSP market capitalization of firm i in year t. Other variables are constructed in the same way as described in Section 3. In this specification, I include a set of firms' log market capitalizations by varying the polynomial order N to control for firms' sizes. In the second stage regression, I estimate the effect of instrumented passive fund holdings from Equation (6) on various measures of price efficiency.

$$\begin{aligned} \mathit{Efficiency}_{i,t} = \beta \widehat{\mathit{Passive}}\%_{i,t} + \sum_{n=1}^{N} \gamma_n (\log(\mathit{Market\ Cap}_{i,t}))^n \\ + \gamma_{N+1} \mathit{FloatAdj}_{i,t} + \kappa_t + \eta_i + \epsilon_{i,t}. \end{aligned} \tag{7}$$

Similar to my main analyses in Section 3, I include both firm and year fixed effects, and all standard errors are clustered at the firm level.

I find that my main finding on the impact of passive investment on stock price efficiency is robust to not only the specification of Equations (6) and (7) but also all to different values for the polynomial N. Table 11 reports the estimated coefficients on $\mathit{Passive\%}_{i,t}$ from the second-stage regression for the pricing error of Hasbrouck (1993) (columns (1), (2), and (3)), the absolute value of return autocorrelation (columns (4), (5), and (6)), and the price delay measure of Hou and Moskowitz (2005) (columns (7), (8), and (9)). I find a statistically negative relationship between $\mathit{Passive\%}_{i,t}$ and all efficiency measures that is robust to various polynomial order controls for market capitalization, indicating the improvement of price efficiency due to an exogenous increase of passive investment.

I further confirm the finding that passive and active investment play a complementary role in the improvement of stock price efficiency. Using the alternative specifications with Equations (6) and (7), I estimate the impact of passive investment on stock price efficiency depending on the investment of active mutual funds. Each panel in Table 12 provides the regression results for each measure of price efficiency. Consistent with findings in Section 5.2, I find that the strongest improvement in stock price efficiency when actively managed funds own significant amounts of shares outstanding. Whereas I observe an insignificant change in price efficiency when active funds own small amounts of shares outstanding.

6.2 Different Bandwidths

In Section 3.3, I discuss the choice of optimal bandwidth around the threshold of the Russell indexes. Recent research using the empirical setting of the Russell indexes use different values of bandwidth. For example, Appel, Gormley, and Keim (2016) use a bandwidth of 250 around the threshold in their main analysis, while Chang, Hong, and Liskovich (2015) use a bandwidth of 100. I choose a bandwidth of 250 in the main analyses based on the investigation on optimal bandwidth using a procedure prescribed in Calonico, Cattaneo, and Titiunik (2014). However, to test the robustness of findings, I reexamine the main results in Section 5 using different bandwidths.

I find that main results of the paper are robust to different choices of bandwidth. Table 13 provides the second-stage estimation results using Equations (3) and (4). Panel A and Panel B report the estimated coefficients for bandwidths of 100 and 500, respectively. In Panel A, I find that the estimated coefficient on $Passive\%_{i,t}$ is statistically significant at the 5% level for the pricing error and the price delay measures, but I do not find statistical significance for the absolute value of autocorrelation. This lack of statistical power in the estimates is due to a small number of observations and a narrow bandwidth to capture variation in passive fund investment after the banding policy of Russell. When I use a bandwidth of 500, I find that, in Panel B, $Passive\%_{i,t}$ is significantly associated with all price efficiency measures at the 1% level.

I also confirm that the main finding on the complementarity of passive and active investment on stock price efficiency is robust to different choices of bandwidth around the index cutoff. Table 14 reports the second-stage regression results using Equations (3) and (4) depending on the investment of active mutual funds. Panel A and Panel B provide the results for bandwidths of 100 and 500, respectively. In Panel A, I do not include the firm-fixed effect due to the limited number of observations for the small bandwidth. Due to limited space, I report the results for the bottom (column (1), (2), (3)) and top (column (4), (5), (6)) quartiles. In unreported results, I also find a monotonic increase in the improvement along with an increase of active investment. Consistent with the finding in Section 5.2, I find the strongest improvement in stock price efficiency when actively managed funds own significant amounts of shares outstanding for both choices of bandwidths.

7 Conclusion

This paper investigates the collective impact of passive and active investment on stock price efficiency. The collaborative nature of objectives of passive and active investment requires the impact of passive and active funds to be jointly analyzed. Using the annual reconstitution of Russell 1000 and 2000 indexes, I document the complementary role of passive and active investment on the discovery of efficient stock prices. For the first set of my findings, I find that an exogenous increase in passive investment improves the efficiency of stock prices. For the second set of my findings, I further find that the improvement of price efficiency arises for stocks that are analyzed and invested by actively managed funds when information is fully shared with passively managed funds.

This paper addresses one of the long standing and important questions in finance regarding the extent to which stock prices reflect information. The impossibility of a perfectly efficient market implies the fact that passively managed funds themselves are not able to make security prices fully efficient because they do not have any incentive to gather and process information. Furthermore, it cannot be an equilibrium where only actively managed funds exist in society. Thus, my finding implies that, combined with the significant presence of actively managed funds which gather and process information, security prices become more efficient as information is fully shared with passively managed funds, and actively managed funds are compensated with high returns for their effort. To the best of my knowledge, the present paper is the only one to investigate the complementary effect of passive and active investment on price efficiency.

A Appendices

A.1 Pricing Error of Hasbrouck (1993)

In this appendix, I explain how to construct one of price efficiency measures used in the paper: the pricing error proposed by Hasbrouck (1993). I follow the procedure and notations prescribed in his paper. Hasbrouck (1993) defines the log transaction price at transaction time t, p_t, as the sum of a random walk component, m_t, and a transitory pricing error, s_t:

$$p_t = m_t + s_t. \tag{8}$$

That is, m_t is defined as the unobservable efficient price or the expected value of the security conditional on all available information at time t whereas the pricing error s_t captures deviations from the efficient price, which may result from non-information-related market frictions such as inventory cost or transaction cost. He proposes the standard deviation of the pricing error, $\sigma(s)$, as a measure of market quality, because this measure captures the magnitude of deviations from the efficient price.

In the empirical estimation, I follow Hasbrouck (1993) in which he estimates the following vector autoregression system with five lags:

$$\begin{aligned} r_t &= a_1 r_{t-1} + a_2 r_{t-2} + \cdots + b_1 x_{t-1} + b_2 x_{t-2} + \cdots + v_{1,t}, \\ x_t &= c_1 r_{t-1} + c_2 r_{t-2} + \cdots + d_1 x_{t-1} + d_2 x_{t-2} + \cdots + v_{2,t}, \end{aligned} \tag{9}$$

where r_t is the difference in the log prices p_t, and x_t is a vector of trade-related variables: a trade sign indicator, signed trading volume, and signed square root of trading volume to allow for concavity between prices and trades. $v_{1,t}$ and $v_{2,t}$ are zero-mean, serially uncorrelated disturbances from the return and the trade equations, respectively. As noted in Hasbrouck (1993), the above VAR system can be converted to its vector moving average (VMA) representation that expresses the variables in terms of contemporaneous and lagged disturbances:

$$\begin{aligned} r_t &= a_0^* v_{1,t} + a_1^* v_{1,t-1} + a_2^* v_{1,t-2} + \cdots + b_0^* v_{2,t} + b_1^* v_{2,t-1} + b_2^* v_{2,t-2} + \cdots \\ x_t &= c_0^* v_{1,t} + c_1^* v_{1,t-1} + c_2^* v_{1,t-2} + \cdots + d_0^* v_{2,t} + d_1^* v_{2,t-1} + d_2^* v_{2,t-2} + \cdots \end{aligned} \tag{10}$$

Using Equation (10) and the identification restriction of Beveridge and Nelson (1981), the pricing error can be expressed as

$$s_t = \alpha_0 v_{1,t} + \alpha_1 v_{1,t-1} + \cdots + \beta_0 v_{2,t} + \beta_1 v_{2,t-1} + \cdots , \tag{11}$$

where $\alpha_j = -\sum_{k=j+1}^{\infty} \alpha_k^*$ and $\beta_j = -\sum_{k=j+1}^{\infty} b_k^*$. Then the variance of pricing error can be computed as

$$\sigma^2(s) = \sum_{j=0}^{\infty} [\alpha_j, \beta_j] Cov(v) \begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix} . \tag{12}$$

In the implementation, I use transaction data in the TAQ database and a filter used in Boehmer and Wu (2013). To assign trade direction, I use the algorithm of Lee and Ready (1991). To assure meaningful analyses, I scale the standard deviation of the pricing error by the standard deviation of log transaction prices $\sigma(p)$. Thus, the ratio of the standard deviation of the pricing errors to that of the efficient price, $\sigma(s)/\sigma(p)$, is referred to as the pricing error in the main analysis.

A.2 Control Variables for Liquidity

In this section, I provide details on how to construct variables used in my main regressions in order to control for liquidity. Goyenko, Holden, and Trzcinka (2009) examine various measures of liquidity and their performance in capturing the price impact, and find that a measure introduced by Amihud (2002) and number of days with positive trading volume and zero returns (zeros) outperform relative to other measures. Thus, in main regressions, I include two measures of liquidity as control variables: an illiquidity measure of Amihud (2002) and zeros defined as the proportion of positive days with zero-returns.

The measure of illiquidity introduced by Amihud (2002) captures the daily price response associated with a dollar of trading volume. Thus, the Amihud measure for firm i in month t is defined as follows:

$$Amihud_{i,t} = \frac{1}{\#of\ Days} \sum_{d=1}^{\#of\ Days} \frac{|R^i_{t,d}|}{V^i_{t,d}}, \tag{13}$$

where $R^i_{t,d}$ and $V^i_{t,d}$ are a return and a dollar trading volume on stock i in day d in month t.

Goyenko, Holden, and Trzcinka (2009) introduce a variable to capture liquidity, defined as the proportion of days (with positive trading volume) with zero returns, as stocks with higher transaction costs have less private information acquisition because it is more difficult to overcome high transaction costs, which leads to have no-information revelation and zero-return days. Thus, the variable, *Zeros*, is constructed as

$$Zeros_{i,t} = \frac{(\#\text{ of positive-volume days with zero returns of stock } i \text{ in month } t\ )}{(\#\text{ of trading days in month } t)}. \tag{14}$$

That is, *Zeros* is the proportion of trading days with positive trading volume and zero return in a given month.

References

Amihud, Yakov, 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

Appel, Ian R., Todd A. Gormley, and Donald B. Keim, 2016, Passive investors, not passive owners, *Journal of Financial Economics* 121, 111–141.

Ball, Ray, and Philip Brown, 1968, An Empirical Evaluation of Accounting Income Numbers, *Journal of Accounting Research* 6, 159–178.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, 2014, Do ETFs Increase Volatility?, NBER Working Papers 20071 National Bureau of Economic Research, Inc.

Bernard, Victor L., and Jacob K. Thomas, 1989, Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium?, *Journal of Accounting Research* 27, 1–36.

Beveridge, Stephen, and Charles R. Nelson, 1981, A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle', *Journal of Monetary Economics* 7, 151–174.

Boehmer, Ekkehart, and Juan (Julie) Wu, 2013, Short Selling and the Price Discovery Process, *Review of Financial Studies* 26, 287–322.

Boone, Audra L., and Joshua T. White, 2015, The effect of institutional ownership on firm transparency and information production, *Journal of Financial Economics* 117, 508–533.

Brav, Alon, Wei Jiang, Frank Partnoy, and Randall Thomas, 2008, Hedge Fund Activism, Corporate Governance, and Firm Performance, *Journal of Finance* 63, 1729–1775.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik, 2014, Robust Nonparametric Confidence Intervals for RegressionDiscontinuity Designs, *Econometrica* 82, 2295–2326.

Cao, Charles, Bing Liang, Andrew W Lo, and Lubomir Petrasek, 2014, Hedge fund holdings and stock market efficiency, *The Review of Asset Pricing Studies* .

Chang, Yen-Cheng, Harrison Hong, and Inessa Liskovich, 2015, Regression Discontinuity and the Price Effects of Stock Market Indexing, *Review of Financial Studies* 28, 212–246.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2005, Evidence on the speed of convergence to market efficiency, *Journal of Financial Economics* 76, 271–292.

Crane, Alan D., Sbastien Michenaud, and James P. Weston, 2016, The Effect of Institutional Ownership on Payout Policy: Evidence from Index Thresholds, *The Review of Financial Studies* 29, 1377.

Da, Zhi, and Sophie Shive, 2016, Exchange traded funds and asset return correlations, *NBER Working Paper* .

Elton, Edwin J., Martin J. Gruber, and Christopher R. Blake, 2011, Holdings Data, Security Returns, and the Selection of Superior Mutual Funds, *Journal of Financial and Quantitative Analysis* 46, 341–367.

Elton, Edwin J, Martin J Gruber, Sanjiv Das, and Matthew Hlavka, 1993, Efficiency with costly information: A reinterpretation of evidence from managed portfolios, *Review of Financial Studies* 6, 1–22.

Falkenstein, Eric G, 1996, Preferences for Stock Characteristics as Revealed by Mutual Fund Portfolio Holdings, *Journal of Finance* 51, 111–135.

Fama, Eugene F., and Kenneth R. French, 2010, Luck versus Skill in the Cross-Section of Mutual Fund Returns, *Journal of Finance* 65, 1915–1947.

Frazzini, Andrea, and Owen A. Lamont, 2008, Dumb money: Mutual fund flows and the cross-section of stock returns, *Journal of Financial Economics* 88, 299–322.

Frino, Alex, and David R Gallagher, 2001, Tracking S&P 500 index funds, *Journal of Portfolio Management* 28, 44–55.

Glosten, Lawrence R, Suresh Nallareddy, and Yuan Zou, 2016, ETF trading and informational efficiency of underlying securities, *SSRN Working Paper* .

Goyenko, Ruslan Y., Craig W. Holden, and Charles A. Trzcinka, 2009, Do liquidity measures measure liquidity?, *Journal of Financial Economics* 92, 153–181.

Griffin, John M., Patrick J. Kelly, and Federico Nardari, 2010, Do Market Efficiency Measures Yield Correct Inferences? A Comparison of Developed and Emerging Markets, *Review of Financial Studies* 23, 3225–3277.

Grossman, Sanford J, and Joseph E Stiglitz, 1980, On the Impossibility of Informationally Efficient Markets, *American Economic Review* 70, 393–408.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw, 2001, Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design, *Econometrica* 69, 201–209.

Hamm, Sophia JW, 2014, The effect of ETFs on stock liquidity, *SSRN Working Paper* .

Hasbrouck, Joel, 1993, Assessing the Quality of a Security Market: A New Approach to Transaction-Cost Measurement, *Review of Financial Studies* 6, 191–212.

Hou, Kewei, and Tobias J. Moskowitz, 2005, Market Frictions, Price Delay, and the Cross-Section of Expected Returns, *Review of Financial Studies* 18, 981–1020.

Iliev, Peter, and Michelle Lowry, 2015, Are Mutual Funds Active Voters?, *Review of Financial Studies* 28, 446–485.

Ippolito, Richard A., 1989, Efficiency with Costly Information: A Study of Mutual Fund Performance, 19651984, *The Quarterly Journal of Economics* 104, 1–23.

Israeli, Doron, Charles M. C. Lee, and Suhas A. Sridharan, 2017, Is there a dark side to exchange traded funds? An information perspective, *Review of Accounting Studies* 22, 1048–1083.

Kokkonen, Joni, and Matti Suominen, 2015, Hedge Funds and Stock Market Efficiency, *Management Science* 61, 2890–2904.

Lee, Charles M C, and Mark J Ready, 1991, Inferring Trade Direction from Intraday Data, *Journal of Finance* 46, 733–746.

Lee, David S., and Thomas Lemieux, 2010, Regression Discontinuity Designs in Economics, *Journal of Economic Literature* 48, 281–355.

Pástor, Luboš, and Robert F Stambaugh, 2012, On the Size of the Active Management Industry, *Journal of Political Economy* 120, 740–781.

Saffi, Pedro A. C., and Kari Sigurdsson, 2011, Price Efficiency and Short Selling, *Review of Financial Studies* 24, 821–852.

Stock, James H, and Motohiro Yogo, 2005, Testing for Weak Instruments in Linear IV Regression, *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg* 80.

Figure 1: Index Portfolio Weights around the Russell 1000/2000 Cutoff in 2006

The figure plots the portfolio weights of the firms around the cutoff for the Russell 1000 and 2000 indexes (bandwidth=250, i.e., bottom 250 firms in the Russell 1000 index and top 250 firms in the Russell 2000 index). The portfolio weights are measured in percentage and plotted against the end-of-June rank of weights in the indexes as of the end of June in 2006.

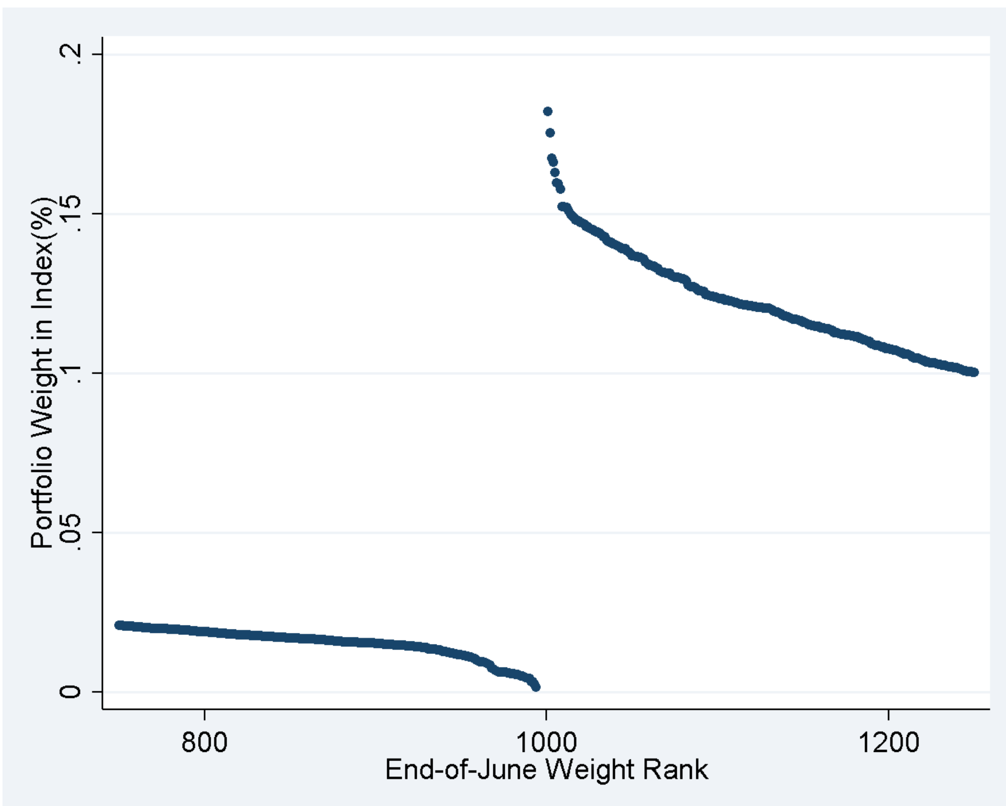

Figure 2: Mutual Fund Holdings Discontinuity around Russell Cutoff

The figure plots the mutual fund holdings after the reconstitution of the Russell 1000 and 2000 indexes from 1996-2016. The top, middle, and bottom graphs represent total, passive, and active mutual fund holdings, respectively. The x-axis represents the rank of weight in the index. Thus, the firms that are in Russell 1000 are on the left-hand side of the horizontal line, and the firms that are in Russell 2000 are on the right-hand side of the line. The y-axis represents the ratio of shares held by mutual funds to total shares outstanding. The figures plot the average mutual fund holdings over 100 bins across all years. The solid line represents a third-order polynomial regression curve.

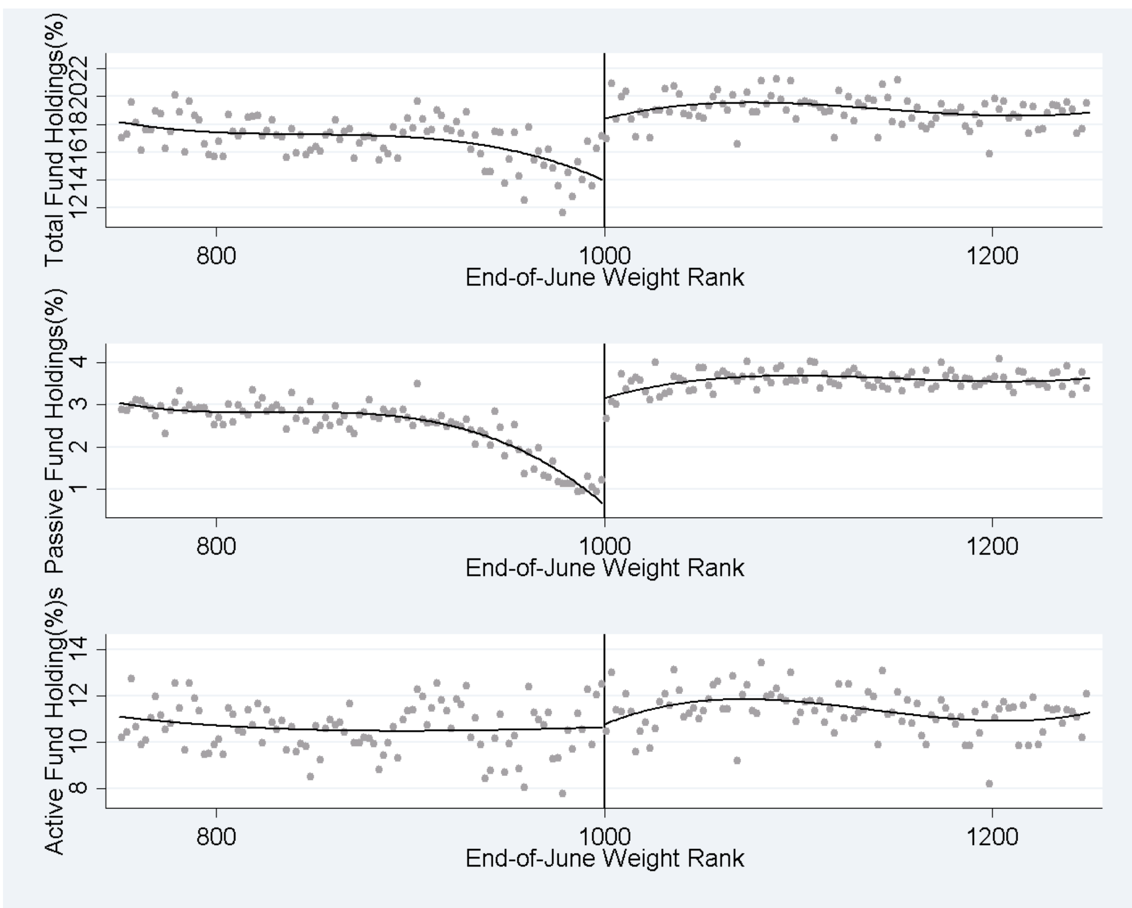

Table 1: Descriptive Statistics

The table reports descriptive statistics of variables of main interest in the paper. Panel A (Panel B) provide mean, 25th percentile, median, 75th percentile, standard deviation, and a number of observations in each column for all firms included in the Russell 1000 and 2000 indexes (firms in a 250 bandwidth around the cutoff between the Russell 1000 and 2000 indexes) from 1997 to 2016. *Total Mutual Fund Holdings*, *Passive Fund Holdings*, *Active Fund Holdings*, and *Unclassified Holdings* are the percentage of share owned by all mutual funds, passive mutual funds, active mutual funds, and unclassified mutual funds, respectively. Holdings data is for the most recent records (from Thomson Reuters Database) after the annual reconstitution of the Russell indexes. *Market Cap* is the market capitalization (in million) of a firm at the end of June each year. *Pricing Error* is the ratio of standard deviation of the discrepancies between the log transaction price and the efficient price to the standard deviation of the efficient prices based on Hasbrouck (1993). |*AR(30)*| is the absolute value of the thirty-minute return autocorrelation following Boehmer and Wu (2013). *Price Delay* and *Amihud* are a measure of price delay following Hou and Moskowitz (2005) and an illiquidity measure of Amihud (2002), respectively. *Zeros* is the proportion of positive-volume days with zero returns.

	Mean	p25	Median	p75	SD	Obs
Panel A. Full Sample						
Total Mutual Fund Holdings (%)	14.55	6.91	12.88	20.71	9.65	51835
Passive Fund Holdings (%)	2.60	0.43	1.70	3.81	2.71	51835
Active Fund Holdings (%)	8.57	3.45	7.26	12.27	6.46	51835
Unclassified Holdings (%)	3.38	0.34	1.37	4.13	5.07	51835
Market Cap (Million)	4874.52	293.18	786.10	2620.20	18812.30	51835
Pricing Error	0.129	0.631	0.098	0.144	0.083	40244
|AR(30)|	0.26	0.25	0.26	0.27	0.02	40244
Price Delay	0.48	0.23	0.44	0.73	0.29	51835
Amihud($\times$100)	1.92	0.07	0.31	1.37	5.09	51835
Zeros	0.03	0.00	0.00	0.05	0.06	51835
Analyst Following (3-month)	8.44	3.00	6.00	11.00	8.45	51835
Analyst Forecast Dispersion (3-month)	0.20	0.02	0.04	0.12	0.56	51835
Panel B. Bandwidth = 250						
Total Mutual Fund Holdings (%)	17.63	9.42	16.33	24.48	10.33	9084
Passive Fund Holdings (%)	2.89	0.32	1.88	4.17	3.12	9084
Active Fund Holdings (%)	10.73	5.50	9.47	14.82	6.92	9084
Unclassified Holdings (%)	4.01	0.61	1.83	4.68	5.56	9084
Market Cap (Million)	1664.13	1142.74	1546.55	2072.99	720.40	9084
Pricing Error	0.936	0.569	0.908	1.549	0.107	6756
|AR(30)|	0.26	0.25	0.26	0.27	0.02	6756
Price Delay	0.46	0.21	0.41	0.70	0.29	9084
Amihud($\times$100)	0.35	0.07	0.14	0.34	1.06	9084
Zeros	0.03	0.00	0.00	0.05	0.05	9084
Analyst Following (3-month)	9.09	4.00	7.00	12.00	7.57	9084
Analyst Forecast Dispersion (3-month)	0.18	0.02	0.04	0.10	0.54	9084

Table 2: Mutual Fund Investment and Price Efficiency around the Index Cutoff

The table reports descriptive statistics of key variables around the Russell index cutoff for different bandwidths depending on the index assignment. Panel A, Panel B, and Panel C reports statistics (mean, median, and standard deviation) of main variables for firms in the 100, 250, and 500 bandwidths, respectively, around the cutoff between the Russell 1000 and 2000 indexes from 1996 to 2016. In each panel, descriptive statistics are reported separately depending on whether firms are assigned in the Russell 1000 index or Russell 2000 index.

Panel A. Bandwidth = 100

	Russell 1000 bottom 100			Russell 2000 top 100		
	Mean	Median	SD	Mean	Median	SD
Total Mutual Fund Holdings (%)	15.66	13.46	9.43	19.25	18.25	11.10
Passive Fund Holdings (%)	1.86	0.94	2.35	3.43	2.58	3.51
Active Fund Holdings (%)	10.23	8.78	6.41	11.59	10.45	7.28
Unclassified Holdings (%)	3.57	1.50	5.09	4.24	1.95	6.01
Market Cap (Million)	1240.32	1114.11	625.67	1854.22	1717.17	652.80
Pricing Error	0.086	0.069	0.112	0.089	0.077	0.121
$\lvert AR(30)\rvert$	0.26	0.26	0.02	0.26	0.26	0.01
Price Delay	0.52	0.49	0.30	0.43	0.36	0.29
Amihud($\times$100)	0.53	0.17	1.51	0.27	0.11	0.48
Zeros	0.03	0.00	0.05	0.03	0.00	0.05

Panel B. Bandwidth = 250

	Russell 1000 bottom 250			Russell 2000 top 250		
	Mean	Median	SD	Mean	Median	SD
Total Mutual Fund Holdings (%)	16.46	15.07	9.52	18.78	17.84	10.95
Passive Fund Holdings (%)	2.36	1.37	2.66	3.40	2.56	3.43
Active Fund Holdings (%)	10.30	9.06	6.61	11.16	9.97	7.19
Unclassified Holdings (%)	3.80	1.80	5.08	4.22	1.87	5.99
Market Cap (Million)	1784.44	1707.17	821.77	1545.62	1453.86	580.37
Pricing Error	0.088	0.713	0.093	0.094	0.068	0.114
$\lvert AR(30)\rvert$	0.26	0.26	0.02	0.26	0.26	0.01
Price Delay	0.49	0.44	0.30	0.44	0.38	0.29
Amihud($\times$100)	0.35	0.13	1.01	0.36	0.15	1.11
Zeros	0.03	0.00	0.05	0.03	0.00	0.05

Panel C. Bandwidth = 500

	Russell 1000 bottom 500			Russell 2000 top 500		
	Mean	Median	SD	Mean	Median	SD
Total Mutual Fund Holdings (%)	16.78	15.60	9.44	18.07	17.10	10.76
Passive Fund Holdings (%)	2.57	1.56	2.79	3.29	2.48	3.31
Active Fund Holdings (%)	10.35	9.22	6.48	10.69	9.51	7.15
Unclassified Holdings (%)	3.86	1.82	5.15	4.09	1.81	5.86
Market Cap (Million)	2565.70	2357.68	1288.22	1250.56	1138.14	548.49
Pricing Error	0.069	0.079	0.068	0.108	0.098	0.124
$\lvert AR(30)\rvert$	0.26	0.26	0.01	0.26	0.26	0.02
Price Delay	0.47	0.42	0.30	0.45	0.39	0.29
Amihud($\times$100)	0.24	0.09	0.78	0.51	0.20	1.38
Zeros	0.03	0.00	0.05	0.03	0.00	0.06

Table 3: Impact of Index Assignment on Mutual Fund Investment

This table reports the regression discontinuity estimates of mutual fund investment on the Russell 1000 and 2000 indexes assignment. Dependent variables are the percentages of share holdings by all mutual funds (column (1)), passive mutual funds (column (2)), active mutual funds (column (3)), and unclassified mutual funds (column (4)). $R2000$ is an indicator variable equal to one if a firm is included in the Russell 2000 index. $Rank^*$ is the rank of a firm based on market capitalization at the time of assignment. c is a cutoff of the Russell 1000 index, which is 1000 before banding policy and is calculated separately every year after banding policy is implemented. $FloatAdj$ is the difference between the rank implied by the end-of-May capitalization and the actual rank in the index assigned by the Russell at the end-of-June. $Amihud$ is the illiquidity measure of Amihud (2002), and $Zeros$ is the proportion of positive-volume days with zero returns. The sample consists of 500 firms around the Russell 1000 and 2000 indexes (i.e., bandwidth = 250) for which I obtain mutual fund holdings data from Thomson Reuters Database. Both year and firm fixed effects are included, and standard errors are clustered at the firm level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Dependent variable: Percentage of holdings by			
	(1)	(2)	(3)	(4)
	All Mutual Funds	Passive	Active	Unclassified
R2000	0.235	0.334***	-0.338	0.203
	(0.60)	(3.77)	(-0.99)	(1.26)
(Rank* - c)	-2.264*	0.208	-1.395	-0.942*
	(-1.84)	(0.69)	(-1.31)	(-1.71)
(Rank* - c) × R2000	-0.516	-0.375***	-0.0302	-0.214*
	(-1.42)	(-4.40)	(-0.10)	(-1.69)
FloatAdj	2.201*	-0.324	1.372	0.917*
	(1.76)	(-1.08)	(1.25)	(1.69)
Amihud	9.896	8.000***	2.198	-1.581
	(1.53)	(5.42)	(0.38)	(-0.92)
Zeros	-6.787***	-0.622*	-4.241***	-1.908***
	(-4.08)	(-1.69)	(-3.12)	(-3.37)
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
R-squared	0.805	0.871	0.692	0.860
Obs	8836	8836	8836	8836

Table 4: Impact of Passive Investment on Price Efficiency

This table reports the results for an instrumental variable estimation of price efficiency on passive mutual fund investment based on Equation (3) and (4). Panel A reports the first stage estimates of passive mutual fund holdings on the Russell index assignment. Panel B reports the estimates of the second stage regression of price efficiency on passive mutual fund holdings estimated from the first stage. *R*2000 is an indicator variable equal to one if a firm is included in the Russell 2000 index. *Pricing Error* is the ratio of standard deviation of the discrepancies between the log transaction price and the efficient price to the standard deviation of the efficient prices based on Hasbrouck (1993). $|AR(30)|$ is the absolute value of the thirty-minute return autocorrelation following Boehmer and Wu (2013). *Price Delay* is a measure of price delay following Hou and Moskowitz (2005). *Rank** is the rank of a firm based on market capitalization at the time of assignment. c is a cutoff of the Russell 1000 index, which is 1000 before banding policy and is calculated separately every year after banding policy is implemented. $FloatAdj$ is the difference between the rank implied by the end-of-May capitalization and the actual rank in the index assigned by the Russell at the end-of-June. *Amihud* is the illiquidity measure of Amihud (2002), and *Zeros* is the proportion of positive-volume days with zero returns. The sample consists of 500 firms around the Russell 1000 and 2000 indexes (i.e., bandwidth = 250) for which I obtain mutual fund holdings data from Thomson Reuters Database. Both year and firm fixed effects are included, and standard errors are clustered at the firm level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. First-stage			
	(1)	(2)	(3)
	Passive(%)	Passive(%)	Passive(%)
R2000	0.498***	0.498***	0.326***
	(8.11)	(8.11)	(6.24)

Panel B. Second-stage			
	(1)	(2)	(3)
	Pricing Error	|AR(30)|	Price Delay
Passive(%)	-0.699***	-0.490***	-1.081***
	(-3.75)	(-3.57)	(-4.49)
(Rank* - c)	0.437*	0.238	0.908***
	(1.81)	(1.14)	(3.41)
(Rank* - c) × R2000	0.00821	0.00142	-0.135*
	(0.07)	(0.02)	(-1.87)
FloatAdj	-0.109	-0.123	-0.899***
	(-0.16)	(-0.51)	(-3.08)
Amihud	-0.495***	-0.217**	0.0411
	(-3.03)	(-2.11)	(1.03)
Zeros	0.752***	0.0375**	0.0644***
	(4.01)	(2.07)	(3.15)
Year FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
R-squared	0.786	0.884	0.114
Obs	6246	6246	8836

Table 5: Double-sorting on Passive and Active Investment and Price Efficiency

The table reports the results of a double-sorting analysis on passive and active investment. Firms in the sample are sorted into terciles based on the percentage shares of passive mutual fund holdings each year. For each tercile, firms are sorted into terciles based on the percentage of active mutual fund holdings. Panel A, Panel B, and Panel C report the results for the pricing error of Hasbrouck (1993), the price delay measure of Hou and Moskowitz (2005), and the absolute value of return autocorrelation of Boehmer and Wu (2013), respectively. *Low* and *High* represent the averages of price efficiency measures for bottom and top terciles, respectively. *Diff(H-L)* provides the differences of price efficiency measures between top and bottom tercile portfolios. T-statistics are provided in the parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Pricing error			
	Passive (%)		
Active (%)	Low	High	Diff(H-L)
Low	0.1692	0.1154	-0.0538***
t-stat			(-4.31)
High	0.1317	0.1029	-0.0288***
t-stat			(-3.47)
Diff(H-L)	-0.0375***	-0.0125**	
t-stat	(-3.76)	(-2.32)	
Panel B. Absolute value of autocorrelation			
	Passive (%)		
Active (%)	Low	High	Diff(H-L)
Low	0.2637	0.2622	-0.0015*
t-stat			(-1.71)
High	0.2634	0.2612	-0.0022*
t-stat			(-1.70)
Diff(H-L)	-0.0003	-0.0010	
t-stat	(-0.73)	(-0.41)	
Panel C. Price delay			
	Passive (%)		
Active (%)	Low	High	Diff(H-L)
Low	0.5552	0.4557	-0.0996***
t-stat			(-3.98)
High	0.5069	0.4305	-0.0764***
t-stat			(-3.27)
Diff(H-L)	-0.0483***	-0.0252**	
t-stat	(-2.99)	(-2.25)	

Table 6: Collective Impact of Passive and Active Investment on Price Efficiency

The table reports the regression results of price efficiency on the passive fund holdings depending on its shares percentage owned by active mutual funds. Panel A, Panel B, and Panel C report the results for the pricing error of Hasbrouck (1993), the price delay measure of Hou and Moskowitz (2005), and the absolute value of return autocorrelation of Boehmer and Wu (2013), respectively. Each column corresponds to the results for quartiles based on the percentage of shares held by active mutual funds. *Rank** is the rank of a firm based on market capitalization at the time of assignment. *c* is a cutoff of the Russell 1000 index, which is 1000 before banding policy and is calculated separately every year after banding policy is implemented. *FloatAdj* is the difference between the rank implied by the end-of-May capitalization and the actual rank assigned in the index by the Russell at the end-of-June. *Amihud* is the illiquidity measure of Amihud (2002), and *Zeros* is the proportion of positive-volume days with zero returns. The sample consists of 500 firms around the Russell 1000 and 2000 indexes (i.e. bandwidth = 250) for which I obtain mutual fund holdings data from Thomson Reuters Database. Both year and firm fixed effects are included, and standard errors are clustered at the firm level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Pricing Error				
	Active Fund Holdings (%)			
	1 (Low)	2	3	4 (High)
Passive(%)	0.189	-0.344	-0.437	-0.891***
	(0.35)	(-1.11)	(-1.71)	(-3.94)
(Rank* - c)	-0.431	0.163	0.099	0.311
	(-1.02)	(0.78)	(0.81)	(-1.35)
(Rank* - c) × R2000	0.081	0.109	-0.0719	-0.0493
	(0.93)	(1.09)	(-0.87)	(-0.56)
FloatAdj	-0.439	-0.761	-0.196	-0.771
	(-0.49)	(-0.88)	(-0.26)	(-0.95)
Amihud	-2.493***	-2.221***	-1.064**	-1.121**
	(-3.18)	(-2.99)	(-2.14)	(-2.22)
Zeros	0.0741	0.0738	0.643	0.431
	(1.01)	(1.12)	(0.89)	(0.69)
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
R-squared	0.822	0.694	0.751	0.843
Obs	1298	1244	1244	1305

Panel B. Absolute Value of Autocorrelation				
	Active Fund Holdings (%)			
	1 (Low)	2	3	4 (High)
Passive(%)	0.215	-0.766	-0.597**	-0.657***
	(0.67)	(-1.07)	(-2.31)	(-2.78)
(Rank* - c)	-0.801	0.530	0.610	0.321
	(-1.04)	(0.45)	(1.16)	(0.89)
(Rank* - c) × R2000	-0.154	0.129	-0.0767	-0.130
	(-0.69)	(0.42)	(-0.48)	(-0.76)
FloatAdj	1.184	-0.494	-0.563	-0.0786
	(1.38)	(-0.37)	(-0.99)	(-0.22)
Amihud	-1.395***	-0.194***	-1.742**	-0.764
	(-4.06)	(-3.01)	(-2.20)	(-1.33)
Zeros	0.0545	0.0377	0.00923	-0.00557
	(0.74)	(0.70)	(0.22)	(-0.18)
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
R-squared	0.902	0.877	0.905	0.891
Obs	1298	1244	1244	1305

Panel C. Price Delay				
	Active Fund Holdings (%)			
	1 (Low)	2	3	4 (High)
Passive(%)	-0.155	-2.381	-0.595*	-1.214***
	(-0.40)	(-0.63)	(-1.66)	(-3.44)
(Rank* - c)	0.517	2.444	0.804	1.016**
	(0.83)	(0.58)	(1.56)	(2.27)
(Rank* - c) × R2000	-0.201	0.0960	-0.161	-0.346**
	(-0.95)	(0.15)	(-1.12)	(-2.26)
FloatAdj	-0.437	-2.481	-0.828	-0.808*
	(-0.71)	(-0.54)	(-1.51)	(-1.80)
Amihud	-0.0271	0.0395	-0.714	-0.347
	(-0.16)	(0.31)	(-1.08)	(-0.49)
Zeros	0.0341	-0.0359	0.0924***	0.0568
	(0.58)	(-0.26)	(2.63)	(1.52)
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
R-squared	0.505	-1.083	0.501	0.217
Obs	1838	1729	1728	1842

Table 7: Performance of Stocks Held by Active Mutual Funds

The table reports the regression results of stock return and trading volume on the passive fund holdings depending on its shares percentage owned by active mutual funds. In Panel A (Panel B), a dependent variable is the cumulative return (the cumulative trading volume) of a stock from July to September, which corresponds to the period from the index reconstitution to the mutual fund holdings report date. Each column corresponds to the results for quartiles based on the percentage of share held by active mutual funds. $Rank^*$ is the rank of a firm based on market capitalization at the time of assignment. c is a cutoff of the Russell 1000 index, which is 1000 before banding policy and is calculated separately every year after banding policy is implemented. *FloatAdj* is the difference between the rank implied by the end-of-May capitalization and the actual rank assigned in the index by the Russell at the end-of-June. Proxies for liquidity (the illiquidity measure of Amihud (2002) and the *Zeros*) are included in the regression. The sample consists of 500 firms around the Russell 1000 and 2000 indexes (i.e. bandwidth = 250) for which I obtain mutual fund holdings data from Thomson Reuters Database. Firm fixed effect is included, and standard errors are clustered at the firm level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Return from July-to-September

	Active Fund Holdings (%)			
	1 (Low)	2	3	4 (High)
Passive(%)	2.483	3.394	4.415*	5.954***
	(0.89)	(1.57)	(1.85)	(2.99)
(Rank* - c)	-0.544	4.980	8.193*	14.35***
	(-0.06)	(0.91)	(1.78)	(4.02)
(Rank* - c) × R2000	1.257	-5.270	-1.138	-1.249
	(0.45)	(-1.50)	(-0.38)	(-0.49)
FloatAdj	0.258	-0.0973	-2.025	-6.129*
	(0.03)	(-0.02)	(-0.46)	(-1.86)
Liquidity Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
R-squared	0.397	0.344	0.335	0.307
Obs	1838	1729	1728	1842

Panel B. Trading Volume from July-to-September

	Active Fund Holdings (%)			
	1 (Low)	2	3	4 (High)
Passive(%)	0.574***	0.736***	0.449***	0.145**
	(5.10)	(4.49)	(6.46)	(2.44)
(Rank* - c)	-1.893***	-1.920***	-1.620***	-1.480***
	(-6.70)	(-9.48)	(-13.72)	(-14.12)
(Rank* - c) × R2000	0.103	0.0243	0.175**	0.101
	(0.95)	(0.20)	(2.07)	(1.31)
FloatAdj	0.912***	1.213***	0.698***	0.440***
	(3.36)	(6.30)	(5.51)	(4.55)
Liquidity Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
R-squared	0.873	0.817	0.853	0.863
Obs	1838	1729	1728	1842

Table 8: Earnings Surprise and Post-Earnings Announcement Drifts

The table reports averages of post-earnings announcement drift of firms sorted on earnings surprises. Stocks (firms in the Russell 1000 and Russell 2000 indexes under the columns *Full Sample* and firms around the bandwidth of 250 under the columns *Bandwidth=250*) in my sample are sorted into quartiles each quarter based on earnings surprises calculated as the difference between actual earnings and the most recent consensus forecast recorded in I/B/E/S. Columns under *Announcement Day* provide abnormal returns on the first trading day after earnings announcement. Columns *[+1, +5]* and *[+1, +10]* report cumulative abnormal returns over 5 and 10 trading days after the announcement, respectively. An abnormal return is calculated as the difference between a raw return and a value-weighted CRSP stock return. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

			Full Sample			Bandwidth =250		
	Quartiles		Announcement Day	[+1, +5]	[+1, +10]	Announcement Day	[+1, +5]	[+1, +10]
1	Most Negative	CAR	-0.313%	-0.543%	-0.574%	-0.426%	-0.618%	-0.745%
		t-stat	(-16.659)	(-6.807)	(-3.735)	(-27.989)	(-9.193)	(-6.348)
2		CAR	-0.196%	-0.273%	-0.219%	-0.301%	-0.414%	-0.415%
		t-stat	(-13.321)	(-4.175)	(-1.332)	(-28.417)	(-8.448)	(-4.191)
3		CAR	0.160%	0.326%	0.395%	0.292%	0.404%	0.443%
		t-stat	(11.003)	(5.001)	(3.527)	(26.940)	(7.156)	(4.397)
4	Most Positive	CAR	0.312%	0.517%	0.589%	0.492%	0.613%	0.647%
		t-stat	(19.418)	(7.745)	(4.763)	(36.904)	(10.562)	(5.859)

Table 9: Passive, Active Investments, and Post-Earnings Announcement Drifts

The table reports averages of post-earnings announcement drift of firms sorted on earnings surprises and active mutual fund investments, depending on the index assignment of a firm. In Panel A, the sample consists of 500 firms around the Russell 1000 and 2000 indexes with the bandwidth of 250. Firms are sorted into quartiles based on earnings surprises calculated as the difference between actual earnings and the most recent consensus forecast recorded in I/B/E/S. The last three columns report the averages of cumulative returns over 5 trading days from the second trading day after the announcement for firms assigned to bottom 250 of the Russell 1000 index, for firms assigned to top 250 of the Russell 2000 index, and the differences of them. Panel B and Panel C report the results for the firms included in the bottom quartile and top quartile of active mutual fund holdings, respectively. An abnormal return is calculated as the difference between a raw return and a value-weighted CRSP stock return. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Firms Around the Cutoff (Bandwidth = 250)

Quartiles			(1) Russell 1000 bottom 250	(2) Russell 2000 top 250	(1)-(2) Diff.
1	Most Negative	CAR	-1.004%	-0.335%	-0.669%***
		t-stat			(4.544)
4	Most Positive	CAR	0.668%	0.460%	0.208%***
		t-stat			(2.760)

Panel B. Low Active Holdings

Quartiles			(1) Russell 1000 bottom 250	(2) Russell 2000 top 250	(1)-(2) Diff.
1	Most Negative	CAR	-2.301%	-0.828%	-1.473%***
		t-stat			(9.825)
4	Most Positive	CAR	1.406%	-0.894%	2.300%***
		t-stat			(11.232)

Panel C. High Active Holdings

Quartiles			(1) Russell 1000 bottom 250	(2) Russell 2000 top 250	(1)-(2) Diff.
1	Most Negative	CAR	-0.811%	0.037%	-0.847%***
		t-stat			(6.321)
4	Most Positive	CAR	0.714%	0.887%	-0.173%**
		t-stat			(2.169)

Table 10: Economic Channels of Efficiency Improvement: Analyst Following and Analyst Forecast Dispersion

The table reports the regression results of the number of analyst following and the analyst forecast dispersion on the passive fund holdings depending on its shares percentage owned by active mutual funds. In Panel A (Panel B), a dependent variable is the the number of unique analyst following (the analyst earnings forecast dispersion) of a stock from July to September, which corresponds to the period from the index reconstitution to the mutual fund holdings report date. Each column corresponds to the results for quartiles based on the percentage of share held by active mutual funds. $Rank^*$ is the rank of a firm based on market capitalization at the time of assignment. c is a cutoff of the Russell 1000 index, which is 1000 before banding policy and is calculated separately every year after banding policy is implemented. $FloatAdj$ is the difference between the rank implied by the end-of-May capitalization and the actual rank assigned in the index by the Russell at the end-of-June. Proxies for liquidity (the illiquidity measure of Amihud (2002) and the *Zeros*) are included in the regression. The sample consists of 500 firms around the Russell 1000 and 2000 indexes (i.e. bandwidth = 250) for which I obtain mutual fund holdings data from Thomson Reuters Database. Firm fixed effect is included, and standard errors are clustered at the firm level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Analyst Following from July-to-September				
	Active Fund Holdings (%)			
	1 (Low)	2	3	4 (High)
Passive(%)	-0.0421	-0.169	0.357	1.336**
	(-0.06)	(-0.26)	(0.48)	(2.11)
(Rank* - c)	-7.236***	-8.194***	-7.216***	-7.654***
	(-6.27)	(-6.88)	(-5.84)	(-6.61)
(Rank* - c) × R2000	1.259**	1.823**	1.241	0.741
	(2.30)	(2.16)	(1.49)	(1.10)
FloatAdj	4.847***	5.956***	4.851***	5.492***
	(4.42)	(5.48)	(4.19)	(5.28)
Liquidity Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
R-squared	0.752	0.787	0.770	0.735
Obs	1666	1639	1690	1796
Panel B. Analyst Forecast Dispersion from July-to-September				
Passive(%)	0.110	-0.0582	-0.0951	-0.387**
	(1.07)	(-0.36)	(-0.55)	(-2.09)
(Rank* - c)	-0.235	0.305	-0.0361	-0.0703
	(-0.72)	(1.10)	(-0.13)	(-0.35)
(Rank* - c) × R2000	0.0831	-0.0725	0.196	-0.00836
	(0.62)	(-0.31)	(1.19)	(-0.05)
FloatAdj	-0.0115	-0.428*	0.0512	0.0330
	(-0.03)	(-1.90)	(0.14)	(0.17)
Liquidity Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
R-squared	0.336	0.489	0.340	0.425
Obs	1412	1457	1545	1650

Table 11: Robustness of Finding to Alternative Specifications: Passive Investment and Price Efficiency

The table shows the robustness of the finding on the impact of passive investment on price efficiency using alternative specification. The regression estimates are obtained from the instrumental variable estimation using Equations (6) and (7). *Pricing Error* is the pricing error based on Hasbrouck (1993). $|AR(30)|$ is the absolute value of the thirty-minute return autocorrelation. *Price Delay* is the measure of price delay based on Hou and Moskowitz (2005). The sample consists of 500 firms around the Russell 1000 and 2000 indexes (i.e., bandwidth = 250) for which I obtain mutual fund holdings data from Thomson Reuters Database. Both year and firm fixed effects are included, and standard errors are clustered at the firm level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Pricing Error			|AR(30)|			Price Delay		
Passive(%)	-0.736***	-0.699***	-0.703***	-0.371***	-0.368***	-0.366***	-0.426***	-0.426***	-0.418***
	(-4.15)	(-4.19)	(-4.24)	(-4.68)	(-4.82)	(-4.84)	(-4.48)	(-4.45)	(-4.42)
Polynomial Order	1	2	3	1	2	3	1	2	3
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.771	0.849	0.898	0.895	0.896	0.896	0.426	0.426	0.428
Obs	5746	5746	5746	5746	5746	5746	8337	8337	8337

Table 12: Robustness of Finding to Alternative Specifications: Passive/Active Investment and Price Efficiency

The table shows the robustness of the finding on complementarity of passive and active investment on price efficiency using alternative specification. The regression estimates are obtained from the instrumental variable estimation using Equations (6) and (7). The estimation is conducted for each quartile sorted each year on the percentage of share held by active mutual fund and for each measure of stock price efficiency. Panel A, Panel B, and Panel C report the results for *Pricing Error* (the pricing error based on Hasbrouck (1993)), $|AR(30)|$ (the absolute value of the thirty-minute return autocorrelation), and *Price Delay* (the measure of price delay based on Hou and Moskowitz (2005)). The sample consists of 500 firms around the Russell 1000 and 2000 indexes (i.e., bandwidth = 250) for which I obtain mutual fund holdings data from Thomson Reuters Database. Both year and firm fixed effects are included, and standard errors are clustered at the firm level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Quartile Portfolios on Active Investment and Pricing Error

	Active Fund Holdings (%)											
	Bottom			2			3			Top		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Passive(%)	0.185	0.188	0.187	-0.211	-0.208	-0.211	-0.379**	-0.379**	-0.378**	-0.845***	-0.846***	-0.912***
	(1.27)	(1.28)	(1.27)	(-1.52)	(-1.48)	(-1.51)	(-2.59)	(-2.59)	(-2.59)	(-4.11)	(-4.13)	(-4.11)
Polynomial Order	1	2	3	1	2	3	1	2	3	1	2	3
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.894	0.893	0.893	0.914	0.915	0.915	0.906	0.905	0.903	0.919	0.920	0.920
Obs	1175	1175	1175	1115	1115	1115	1110	1110	1110	1187	1187	1187

Panel B. Quartile Portfolios on Active Investment and Absolute Autocorrelation

	Bottom			2			3			Top		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Passive(%)	-0.223	-0.239*	-0.238*	-0.490***	-0.485***	-0.489***	-0.456***	-0.415**	-0.415**	-0.735**	-0.748***	-0.806***
	(-1.58)	(-1.70)	(-1.74)	(-3.38)	(-3.33)	(-3.33)	(-2.74)	(-2.57)	(-2.57)	(-2.48)	(-2.66)	(-2.68)
Polynomial Order	1	2	3	1	2	3	1	2	3	1	2	3
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.894	0.893	0.893	0.914	0.915	0.915	0.906	0.905	0.903	0.919	0.920	0.920
Obs	1175	1175	1175	1115	1115	1115	1110	1110	1110	1187	1187	1187

Panel C. Quartile Portfolios on Active Investment and Price Delay

	Bottom			2			3			Top		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Passive(%)	-0.214	-0.221	-0.207	-0.318**	-0.319**	-0.327**	-0.547***	-0.541***	-0.547***	-1.805***	-1.785***	-1.728***
	(-1.12)	(-1.13)	(-1.01)	(-2.05)	(-2.05)	(-2.11)	(-2.69)	(-2.65)	(-2.68)	(-2.70)	(-2.67)	(-2.78)
Polynomial Order	1	2	3	1	2	3	1	2	3	1	2	3
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.482	0.481	0.480	0.596	0.596	0.597	0.0126	0.00104	0.0391	0.468	0.469	0.469
Obs	1718	1718	1718	1596	1596	1596	1599	1599	1599	1715	1715	1715

Table 13: Robustness of Finding to Different Bandwidths: Passive Investment and Price Efficiency

This table reports the results of an instrumental variable estimation of price efficiency on passive mutual fund investment based on Equations (3) and (4) when I use different bandwidths around the Russell indexes cutoff. Panel A and B provide the results when the bandwidths are 100 and 500, respectively. *Pricing Error* is the pricing error measure of Hasbrouck (1993). |*AR(30)*| is the absolute value of the thirty-minute return autocorrelation. *Price Delay* is the measure of price delay following Hou and Moskowitz (2005). *Rank** is the rank of a firm based on market capitalization at the time of assignment. *c* is a cutoff of the Russell 1000 index, which is 1000 before banding policy and is calculated separately every year after banding policy is implemented. *FloatAdj* is the difference between the rank implied by the end-of-May capitalization and the actual rank in the index assigned by the Russell at the end-of-June. *Amihud* is the illiquidity measure of Amihud (2002), and *Zeros* is the proportion of positive-volume days with zero returns. Both year and firm fixed effects are included, and standard errors are clustered at the firm level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Panel A. Small Bandwidth = 100			*Panel B. Large Bandwidth=500*		
	(1)	(2)	(3)	(4)	(5)	(6)
	Pricing Error	|AR(30)|	Price Delay	Pricing Error	|AR(30)|	Price Delay
Passive(%)	-0.342**	0.0109	-0.894**	-0.711***	-0.392***	-0.748***
	(-2.33)	(0.04)	(-2.45)	(-4.87)	(-4.71)	(-6.53)
(Rank* - c)	0.019	-1.021	0.998	0.304**	0.0111	0.481***
	(1.11)	(-1.04)	(1.14)	(2.01)	(0.16)	(6.79)
(Rank* - c) × R2000	0.00551	0.197	-0.506***	0.011	0.0315	-0.193***
	(0.08)	(1.13)	(-3.06)	(0.21)	(0.60)	(-4.31)
FloatAdj	0.963	1.134	-0.999	-0.174	0.136	-0.443***
	(0.65)	(1.07)	(-1.08)	(-0.39)	(1.54)	(-5.23)
Amihud	-0.222*	-0.231**	0.0581*	-0.554***	-0.198***	0.0628*
	(-1.74)	(-2.43)	(1.70)	(-3.98)	(-2.83)	(1.68)
Zeros	0.391***	0.0283	0.0886***	0.899***	0.0334***	0.0669***
	(-2.99)	(0.78)	(2.65)	(4.91)	(2.89)	(5.72)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.844	0.897	0.334	0.229	0.247	0.348
Obs	2175	2175	3048	12954	12954	18390

Table 14: Robustness of Finding to Different Bandwidths: Passive/Active Investment and Price Efficiency

This table reports the results for an instrumental variable estimation of the complementary role of passive and active investment in the efficient price discovery based on Equations (3) and (4) when I use different bandwidths around the Russell indexes cutoff. Panel A and B provide the results when the bandwidths are 100 and 500, respectively. *Pricing Error* is the pricing error measure of Hasbrouck (1993). $|AR(30)|$ is the absolute value of the thirty-minute return autocorrelation. *Price Delay* is the measure of price delay following Hou and Moskowitz (2005). *Rank** is the rank of a firm based on market capitalization at the time of assignment. c is a cutoff of the Russell 1000 index, which is 1000 before banding policy and is calculated separately every year after banding policy is implemented. *FloatAdj* is the difference between the rank implied by the end-of-May capitalization and the actual rank in the index assigned by the Russell at the end-of-June. *Amihud* is the illiquidity measure of Amihud (2002), and *Zeros* is the proportion of positive-volume days with zero returns. Only year fixed effect is included in Panel A, while both year and firm fixed effects are included in Panel B. Standard errors are clustered at the firm level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Small Bandwidth = 100

	Active Fund Holdings (%)					
	Bottom			Top		
	(1)	(2)	(3)	(4)	(5)	(6)
	Pricing Error	|AR(30)|	Price Delay	Pricing Error	|AR(30)|	Price Delay
Passive(%)	-0.0181	-0.0383	-0.256***	-0.249*	-0.254***	-0.654***
	(-0.23)	(-0.57)	(-3.30)	(-1.71)	(-2.73)	(-4.17)
(Rank* - c)	-0.275	-0.805	0.643	-1.046	-0.732	1.353
	(-0.18)	(-0.91)	(0.85)	(-1.46)	(-0.67)	(1.32)
(Rank* - c) × R2000	-0.541**	-0.249	-0.552***	0.175	0.0850	-0.200
	(-2.22)	(-1.63)	(-4.70)	(1.18)	(0.43)	(-1.20)
FloatAdj	0.862	1.227	-0.548	1.249*	0.999	-1.397
	(0.59)	(1.40)	(-0.72)	(1.72)	(0.89)	(-1.35)
Amihud	-1.733***	-1.901***	0.241	-0.210	-0.146	0.397
	(-6.37)	(-4.97)	(1.21)	(-0.44)	(-0.30)	(1.19)
Zeros	0.111	0.0798*	0.127***	0.0116	-0.0199	0.0187
	(1.41)	(1.89)	(3.15)	(0.29)	(-0.36)	(0.40)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.820	0.836	0.292	0.859	0.863	0.0577
Obs	656	656	962	651	651	950

Panel B. Large Bandwidth = 500

	Active Fund Holdings (%)					
	Bottom			Top		
	(1)	(2)	(3)	(4)	(5)	(6)
	Pricing Error	|AR(30)|	Price Delay	Pricing Error	|AR(30)|	Price Delay
Passive(%)	-0.182*	-0.174***	-0.328***	-0.312***	-0.319***	-0.460***
	(-1.95)	(-3.37)	(-5.62)	(-5.61)	(-4.71)	(-5.66)
(Rank* - c)	0.241**	0.0758	0.307***	-0.00733	0.0653	0.356***
	(2.05)	(1.02)	(4.58)	(-0.10)	(0.71)	(4.32)
(Rank* - c) $\times$ R2000	0.0917	0.190**	-0.182***	0.0495	-0.0125	-0.209***
	(0.86)	(2.45)	(-2.84)	(0.70)	(-0.15)	(-3.11)
FloatAdj	0.0353	0.151*	-0.360***	0.204***	0.157	-0.298***
	(0.29)	(1.92)	(-5.34)	(2.71)	(1.63)	(-3.52)
Amihud	-1.040***	-1.150***	0.106	-0.568	-0.404	-0.0727
	(-5.56)	(-5.94)	(1.39)	(-1.21)	(-0.87)	(-0.79)
Zeros	0.129***	0.0909***	0.105***	0.0131	-0.00274	0.0733***
	(4.19)	(4.48)	(6.61)	(0.77)	(-0.13)	(3.53)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.820	0.833	0.223	0.857	0.871	0.177
Obs	3311	3311	4793	3347	3347	4798