ESSAYS IN INFORMATION AND ASSET PRICES

A Thesis
Presented to
The Academic Faculty

by

Youngmin Choi

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
Scheller College of Business

Georgia Institute of Technology
August 2018

ESSAYS IN INFORMATION AND ASSET PRICES

Approved by:

Professor Suzanne S. Lee, Advisor,
Committee Chair
Scheller College of Business
Georgia Institute of Technology

Professor Alex Hsu
Scheller College of Business
Georgia Institute of Technology

Professor Shijie Deng
H. Milton Stewart School of Industrial
and Systems Engineering
Georgia Institute of Technology

Professor Sudheer Chava
Scheller College of Business
Georgia Institute of Technology

Professor Soohun Kim
Scheller College of Business
Georgia Institute of Technology

Date Approved: 17 May 2018

To my parents and Hyunah,

for their endless love, support, and encouragement.

ACKNOWLEDGEMENTS

I want to express my sincere gratitude to my advisor Prof. Dr. Suzanne S. Lee who has been a tremendous mentor for me. I am deeply indebted to Dr. Lee for her unwavering support, guidance, and advice on both research as well as on my career. I also want to express my special appreciation and thanks to Prof. Sudheer Chava and Prof. Soohun Kim for his constant support and encouragement throughout my graduate studies. My thanks also goes to the members of my committee Prof. Alex Hsu and Prof. Shijie Deng for providing many valuable comments that improved the presentation and contents of this dissertation.

I am thankful to all the other finance faculty members at the Scheller College of Business for their role in creating a supportive environment for the students. My thanks also goes to Rohan Rao Ganduri, Steven Chong Xiao, Peter Simasek, Nikhil Paradkar, Linghang Zeng, and other fellow Ph.D. students for many interesting and good-spirited research related discussions and for all the fun we have had over the past years.

Finally, I would like to acknowledge the people who mean the world to me, my parents, and sister. I am greatly indebted to them for showing faith in me, and for all the sacrifices they made which allowed me to pursue what I desired. I dedicate this dissertation to them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This thesis uncover the dynamics of asset prices in response to informational events. These studies can provide insights to understand not only the behavior of financial market participants but also the economic mechanisms by which asset prices form. The first essay, "Complementarity of Passive and Active Investment on Stock Price Efficiency", examines how passive and active investment collectively affect the efficiency of stock prices. I document the complementary role of passive and active investment in the improvement of stock price efficiency. The relationship between stock price efficiency and passive/active investment, which is one of the long-standing theme in finance, is subject to an endogeneity problem. Using the annual reconstitution of Russell 1000 and 2000 indexes as an instrument, I find 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 returns for their effort. An increase in analyst following and a decrease in analyst forecast dispersion are identified as economic channels of the efficiency improvement. The result implies that active funds seek out inefficient stocks and ultimately experience superior returns due to the improvement in efficiency from passive investment. This study highlights the complementarity of passive and active investment on stock price efficiency, which has not been documented in the literature.

In the second essay, "The Role of Efficient Analysts in Stock and Option Markets", we investigate the fundamental role of analyst recommendations in terms of the true efficient price discovery using the signal-to-noise volatility ratio constructed with high-frequency data. While exiting studies on the effect of analyst recommendations focus on their ex post impact on stock prices, we find that only recommendation revisions

that contribute to the true efficient price discovery play an influential role in the stock and option markets. In particular, abnormal stock returns are observed in expected directions only for the recommendations that deliver information about the true efficient price of a stock. Furthermore, we observe that those informative revisions resolve uncertainty about a firm and reduce jump risk in its stock price. This study highlights the intrinsic role of analysts in financial market, and documents that the analysts who do their fundamental work indeed affect stock and option markets.

In the third essay,"Realized Skewness for Information Ambiguity", we propose realized skewness constructed using high-frequency data as a measure of information ambiguity. In this working paper, we rely on the fact that ambiguity-averse investors respond asymmetrically to ambiguous (intangible) information. We find that a significant decrease of realized skewness around analysts' earnings forecasts and recommendation releases. Furthermore, as realized skewness proxies the degree of information ambiguity, we document that negative realized skewness predicts subsequent lower returns around information releases after controlling for return continuations. To further prove an economic significance of our finding, we provide a zero-net investment strategy incorporating our finding, which achieves a Sharpe ratio of 1.766 with 0.83% of monthly average returns. The results highlight how the asymmetric behavior of investors in response to information releases can be used to infer subsequent dynamics of asset prices.

CHAPTER I

COMPLEMENTARITY OF PASSIVE AND ACTIVE INVESTMENT ON STOCK PRICE EFFICIENCY

1.1 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

[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, [11] and [22]. [24] document the impact of hedge funds on corporate governance. Among many others, [61] document a relationship between mutual fund flows and the cross-section of stock returns. [29] and [87] show that hedge funds improves the efficiency of stock prices.

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 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 ([69]). 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 ([96]). 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 ([78], [50]).

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. 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 ([17], [43]). 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,

[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 [29] and [87] among many others.

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, [22]) and improvement in governance quality (see, for example, [11]).

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, [52] 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 [34]). 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 [42]. 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 [73], which captures the temporary deviation of a transaction price from the (unobservable) efficient price of a security. Second, following [20], 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 [74], 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 [12]). 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, [22] 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 ([57], [49]). 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). [71] finds that ETFs and passive mutual funds deprive the liquidity of the underlying security, and [80] show that an increase in ETF ownership is associated with reduced price efficiency. [65] 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 [11]) and information disclosure behavior ([22]). 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 1.2 introduces the various measures of price efficiency used in the main analysis. Section 1.3 explains background information on the construction of the Russell indexes and my empirical design. Section 1.4 describes the data. Section 1.5 provides empirical results. In

Section 1.6, I describe several robustness tests. Section 1.7 concludes.

1.2 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.

1.2.1 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 [73]. 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 [73] and [20], 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 [20] and [38], I choose a thirty-minute interval to estimate return autocorrelation of transaction prices and denote |AR30| as the absolute value of the autocorrelation.

1.2.2 Low-frequency efficiency measure

For a low-frequency measure of price efficiency, I construct a price delay measure introduced by [74], 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. [68] and [101] apply the price delay measure in an international context.

The original price delay measure suggested by [74] 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 [20], I modify the approach of [74] 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:

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

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

1.2.3 Post-earnings announcement drift

Since [12] 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.

1.3 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, [52] 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.

1.3.1 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. [34] 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

[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 1.3.2.

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 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

[8] [34] also computed these implied cutoffs each year to examine the price effects of index inclusion, while [11] and [42] 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.

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 [62]). For example, if a stock is assigned into the top 250 of the Russell 2000 index from the bottom 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 [34] 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.

1.3.2 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 [90]. 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:

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

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 [5] measure and zeros

[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. [34] discuss about reasons why the actual weights or rankings should not be used as an instrument.

introduced by [67]. The construction of these variables are explained in detail in Appendix A.2)[10], as [22] 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 holdings on various measures of price efficiency.

$$\begin{aligned} \mathit{Efficiency}_{i,t} = \beta \widehat{Passive\%}_{i,t} + \gamma_1 (Rank^*_{i,t} - c) + \gamma_2 \mathit{Russell2000}_{i,t} (Rank^*_{i,t} - c) \\ + \gamma_3 \mathit{FloatAdj}_{i,t} + \gamma_4 \mathit{Liquidity\ Controls} + \kappa_t + \eta_i + \epsilon_{i,t}, \end{aligned} \tag{4}$$

where $\mathit{Efficiency}_{i,t}$ 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 $\mathit{Russell2000}_{i,t}(Rank^*_{i,t} - c)$ of firm i in year t.[11] Thus, my key instrument is $\mathit{Russell2000}_{i,t}$, conditional on market capitalization ranking, $(Rank^*_{i,t} - c)$, and the interaction $\mathit{Russell2000}_{i,t}(Rank^*_{i,t} - c)$. Both regressions include year and firm fixed effects. All standard errors from the estimation of the above regressions are clustered at the firm level.

1.3.3 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

[10] [67] examine and run horseraces of various measures of liquidity. They find that the illiquidity measure of [5] and zeros (the proportion of positive-volume days with zero returns) outperform in capturing the price impact.

[11] $(Rank^*_{i,t} - c)$ and $\mathit{Russell2000}_{i,t}(Rank^*_{i,t} - c)$ 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.

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. [34] use a bandwidth of 100 around the cutoff in estimating the deletion and addition effect of indexing. [11] use a bandwidth of 250 in their main analysis on the influence of passive owners, while [42] provide the results 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 [28]. 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 1.6.

1.4 Data

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

Two measures of price efficiency explained in Section 1.2.1 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

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

I/B/E/S. Other accounting variables are obtained from the COMPUSTAT database.

1.4.2 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 [76] and [11]. 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.

1.4.3 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.

1.5 *Empirical Results*

1.5.1 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 1.3.3, 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 [22] 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.

1.5.1.1 *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:

$$\begin{aligned} MF\ Holdings\%_{i,t} &= \tau Russell2000_{i,t} + \delta_1(Rank^*_{i,t} - c) + \delta_2 Russell2000_{i,t}(Rank^*_{i,t} - c) \\ &+ \delta_3 FloatAdj_{i,t} + \delta_4 Liquidity\ Controls \alpha_t + \theta_i + \varepsilon_{i,t}, \end{aligned} \tag{5}$$

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 1.3.2.

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 [5] 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.

1.5.1.2 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 ($Efficiency_{i,t}$) and passive fund holdings percentage ($Passive\%_{i,t}$) 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 $Passive\%_{i,t}$.

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

[13] [11] 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 [42] 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.

[14] For all columns in Table 4, F-statistics and t-statistics exceed the thresholds suggested by [103].

price efficiency measures: the pricing error of [73] (column (1)), the absolute value of autocorrelation (column (2)), and the price delay of [74] (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 1.6.

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

1.5.2 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. [69] 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 [73], the absolute value of autocorrelation, and the price delay of [74], 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 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

[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 [103].

[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, [34] and [42].

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.

1.5.3 Passive/Active investment and post-earnings announcement drift

In this subsection, I analyze how passive investment is associated with post-earnings announcement drift. [12] 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. [18] 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 1.5.1.

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, [22] 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 1.5.2, 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 analyst forecast dispersion with passive fund investment for other lower quartiles.

1.6 Robustness Check

1.6.1 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

[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.

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, [90] and [70]). 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. [34] 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. [11] 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 [11]. In particular, I estimate the following first-stage regression:

$$\begin{aligned} Passive\%_{i,t} = {} & \tau Russell2000_{i,t} + \sum_{n=1}^{N} \delta_n (\log(Market\ Cap_{i,t}))^n \\ & + \delta_{N+1} FloatAdj_{i,t} + \alpha_t + \theta_i + \varepsilon_{i,t}, \end{aligned} \tag{6}$$

where $Market\ Cap_{i,t}$ 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 1.3. 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.

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

Similar to my main analyses in Section 1.3, 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 $Passive\%_{i,t}$ from the second-stage regression for the pricing error of [73] (columns (1), (2), and (3)), the absolute value of return autocorrelation (columns (4), (5), and (6)), and the price delay measure of [74] (columns (7), (8), and (9)). I find a statistically negative relationship between $Passive\%_{i,t}$ 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 1.5.2, 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.

1.6.2 Different Bandwidths

In Section 1.3.3, 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, [11] use a bandwidth of 250 around the threshold in their main analysis, while [34] 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 [28]. However, to test the robustness of findings, I reexamine the main results in Section 1.5 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 1.5.2, I find the strongest improvement in stock price efficiency when actively managed funds own significant amounts of shares outstanding for both choices of bandwidths.

1.7 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.

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.

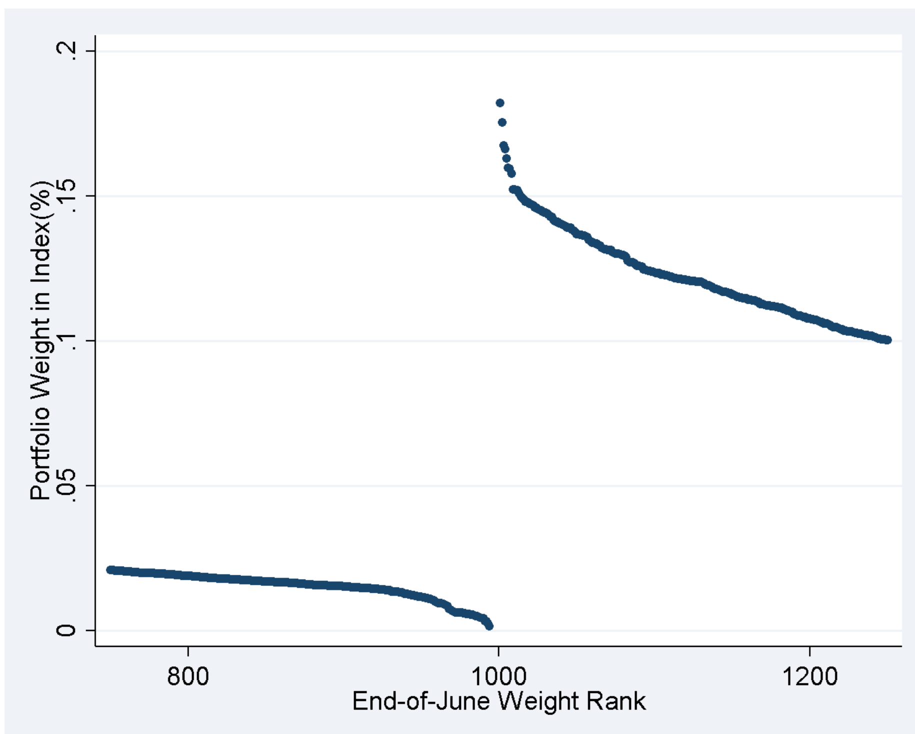


Figure 2: Mutual Fund Holdings Discontinuity around Russell Cutoff

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.

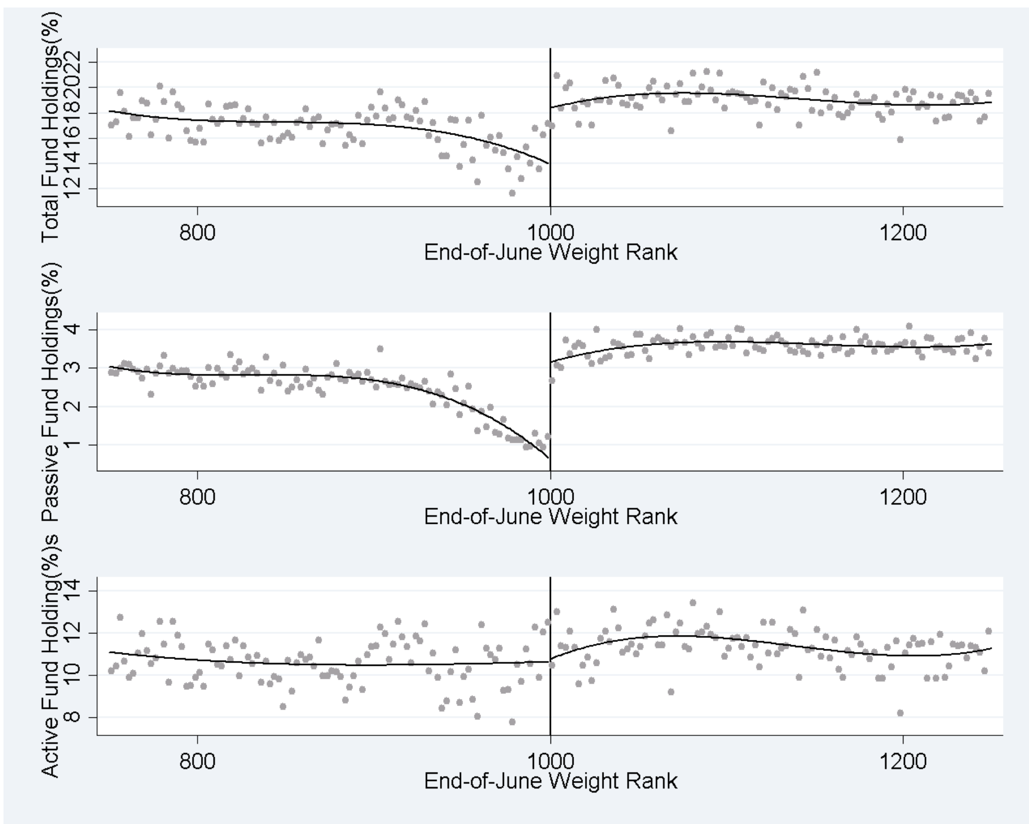


Table 1: Descriptive Statistics

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 [73]. $|AR(30)|$ is the absolute value of the thirty-minute return autocorrelation following [20]. *Price Delay* and *Amihud* are a measure of price delay following [74] and an illiquidity measure of [5], 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
|AR(30)|	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(×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
|AR(30)|	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(×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
|AR(30)|	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(×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 [5], 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. $R2000$ 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 [73]. $|AR(30)|$ is the absolute value of the thirty-minute return autocorrelation following [20]. *Price Delay* is a measure of price delay following [74]. *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 [5], 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 [73], the price delay measure of [74], and the absolute value of return autocorrelation of [20], 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 [73], the price delay measure of [74], and the absolute value of return autocorrelation of [20], 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 [5], 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) $\times$ 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) $\times$ 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 [5] 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 [5] 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 [73]. $|AR(30)|$ is the absolute value of the thirty-minute return autocorrelation. *Price Delay* is the measure of price delay based on [74]. 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 [73]), $|AR(30)|$ (the absolute value of the thirty-minute return autocorrelation), and *Price Delay* (the measure of price delay based on [74]). 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 [73]. $|AR(30)|$ is the absolute value of the thirty-minute return autocorrelation. *Price Delay* is the measure of price delay following [74]. $Rank^*$ 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 [5], 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 [73]. *|AR(30)|* is the absolute value of the thirty-minute return autocorrelation. *Price Delay* is the measure of price delay following [74]. *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 [5], 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) × 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

CHAPTER II

THE ROLE OF EFFICIENT ANALYSTS IN STOCK AND OPTION MARKETS

2.1 Introduction

Security analysts are important contributors to the information efficiency of financial markets. They gather and process a variety of information about a security and provide refined opinions about its attractiveness to other market participants. As noted in [53], their efforts are fundamental to making financial markets more efficient. In other words, it is expected on average that they deliver valuable information through their recommendations, making transaction prices move closer to the intrinsic value of the stock. Thus, the main objective of this paper is to examine the fundamental contribution of analysts to stock price discovery as information intermediaries. In particular, we aim to differentiate the degree of their contribution toward efficient price discovery using individual analyst recommendations rather than the consensus recommendation. We believe understanding this heterogeneity in the analysts' contributions is important because their contribution to price discovery fundamentally justifies the reaction of stock prices.

Existing studies in the literature have investigated the average impact of analyst recommendations by measuring mean (or median) magnitude (i.e., absolute value) of market price reactions or abnormal stock returns to analyst reports.[1] Unlike these studies on the overall impact of analyst recommendations, [92] consider one dimension of heterogeneity across different analysts, which is whether the recommendation

[1]See [59], [60], and [81], among many others.

is influential/noticeable or not. In this paper, we start by examining a different dimension of heterogeneity on how much a recommendation of an analyst contributes to efficient price discovery. Then, we further investigate the impact of recommendations contributing to the efficient price discovery in stock and option markets, where implications for participants' investment decision making and uncertainty resolution due to the contribution of analysts can be well-examined.

In order to assess individual analysts' direct contribution toward efficient price discovery, we measure how much influence their recommendation revisions have on variations in the fundamental value of a corresponding stock. We make the usual assumption that observed prices can be decomposed into the true efficient price (intrinsic value or random walk component) of a security and the noise component (pricing error or market microstructure noise).[2] Then, we measure the variation in efficient prices generated by the recommendation revisions relative to the variation in noise in observed prices. If an analyst delivers an important information contributing to making the relevant security price more efficient, changes in observed prices in response to the information are expected to be driven more by efficient prices than noise component. This idea of using variances to measure the impact of information releases is in a similar spirit of [16], which captures the information content of earnings announcements. Based on this intuition, we propose using the ratio of the volatility of the true efficient price process to the volatility of the noise in transaction prices in order to capture the fundamental role of analysts' recommendation.

Since the approach is based on two nonparametric volatility estimators, our signal-to-noise volatility ratio is robust to model specification. It is also less data intensive compared to other parametric approaches such as [73] who specifies a set of vector autoregressive models to estimate the variance of the true efficient price component and the variance of the noise component. While the pricing error of [73], which is

[2]We follow existing studies for this assumption. Among many others, see [73], [13], and [14].

basically a reciprocal of our SNR measure, has been widely used in various studies, this is the first paper that investigates the impact of analysts' recommendations on price efficiency with our new robust approach.[3]

Using the SNR, we first confirm that analysts on average contribute to efficient price discovery. Specifically, we find evidence that the SNR is significantly associated with both downgrade and upgrade recommendation revisions even after controlling for other firm characteristics as well as firm- and time- fixed effects. After confirming evidence of average impact, we further study the degree of their contribution toward efficient price discovery and their different impact on the stock market. While security analysts make an effort to collect and process information for their analyses, we expect some degree of heterogeneity in their contributions to efficient price discovery. To make the discussion simple, we refer to these analysts contributing to efficient price discovery as "efficient analysts" and the other analysts not contributing to efficient price discovery as "noisy analysts." Effort and resources that analysts spend for their analyses to discover the efficient price justify abnormal returns generated by stock price changes. Consistently, we find evidence that revisions of efficient analysts with higher SNRs generate greater abnormal stock returns than revisions of noisy analysts with lower SNRs. Furthermore, we document that only highly efficient revisions with high SNRs produce significant abnormal stock returns, whereas other revisions do not generate any significant stock price reactions. This finding suggests that analysts contributing to stock market efficiency eventually generate stock price reaction in actual transactions.

A closer analysis further reveals that only those efficient analysts' recommendation revisions generate abnormal stock returns in expected directions. In particular,

[3]Recently, [99] use a variance ratio of 1-minute and 30-minute realized variance estimates as a measure of efficiency. In this paper, we use different frequencies for our measure of efficiency because we believe that 1-second and 5-minute intervals are appropriate choices for our sample firms with the guidance of the empirical analysis of [13] and [14].

only downgrade and upgrade revisions with higher SNRs produce negative and positive abnormal returns around the issuance, respectively. However, noisy analysts' downgrade revisions with lower SNRs generate unexpected positive abnormal returns, possibly due to increase in noise variance. Noisy analysts' upgrade revisions with lower SNRs do not create any significant abnormal returns. This finding implies that only those efficient analysts contributing to stock market efficiency eventually move stock prices in the predicted direction as their recommendations suggest.

In addition to stock market reactions, we study the role of efficient analysts in option markets to examine their impact on the activity of informed traders and the resolution of uncertainty. First, if a recommendation revision delivers valuable information about a firm's fundamentals, we expect committed or informed traders to execute their trades where the potential profit from the transactions can be maximized. Option markets allow us to investigate the association of analysts' recommendation revisions with informed trading, as informed traders utilize options to exploit the informational advantage as discussed in [47] and [32], among others. Recently, [98] suggest that the relative trading activity in option markets, measured by the ratio of option trading volume to the underlying stock trading volume (O/S ratio) is significantly associated with informed trading in option markets. Based on this finding, we hypothesize that a recommendation revision made by efficient analysts would increase trading activity of option markets relative to that of underlying stock markets. Consistent with the hypothesis, we find that the O/S ratio significantly increases when the recommendation revisions with high SNRs are issued. This evidence suggests that option market participants recognize and attempt to exploit the issuance of efficient analysts' recommendation revisions, which deliver valuable information about the intrinsic value of a security.

Second, if a recommendation revision contributes to discover the efficient price of a security, uncertainty about a firm's future prospect can be significantly resolved.

In theory, the release of important information about a firm's performance and sustainability resolves the uncertainty about the firm's future prospects, as shown for earnings announcements in [48]. We test whether efficient analysts play a similar role in relevant markets. Since the price of an option contract reflects market participants' expectation about the firm, we expect that recommendation revisions made by efficient analysts help resolve uncertainty about the related firm and also reduce jump risks in stock prices. Consistent with our expectation, we find evidence that an issuance of downgrade recommendation revisions with high SNRs reduces future uncertainty measured with the average implied volatility of at-the-money (ATM) call and put options of a firm. Furthermore, we also document that downgrade revisions with high SNRs reduce the jump risks in stock prices, which is measured with the slope of implied volatility of out-of-the-money (OTM) put options following [86]. This finding suggests that future stock price movements will be relatively smooth because the issuance of efficient analysts' recommendations works toward resolving extreme uncertainty about a firm.

Our study contributes to several strands of literature in finance. First, to the best of our knowledge, this is the only paper in the literature which differentiates the impact of security analysts from that of noisy analysts on contributing to efficient price discovery. Most existing studies have used average effects (see [81]), influential individual recommendations[4] (see [92]), or stock price reaction (see [60]) to measure their impact. These existing studies implicitly assume that significant reactions observed in stock prices indicate the strong impact of analyst recommendations without considering the presence of noise. Therefore, it is impossible to tell whether the changes in these measures are due to valuable information regarding the intrinsic value of a firm or due to the noise coming from other confounding effects, liquidity issues, or market frictions. Unlike these studies, our proposed SNR measure controls for the

[4] [92] focus on individual recommendations that visibly affect stock price.

amount of noise in the market, and hence robust to their presence while capturing the impact of efficient analysts' recommendations.

Second, this is the first paper which utilizes information from high-frequency data in order to examine the fundamental role of analysts in contributing to efficient price discovery. Despite the advance in the realized variance literature,[5] there have not been many studies which actually take advantage of the development of techniques to address economically critical questions in the literature on analysts, namely distinguishing efficient analysts from noisy analysts. As demonstrated in this paper, the measure can allow us to differentiate their impact in the relevant markets.

Finally, this is the first paper to investigate the impact of efficient analysts on option markets. There is no study particularly focusing on how analysts' recommendation revisions would affect option markets, while existing papers studies options prices in response to scheduled news events such as earnings announcement ([45], [44], [79]), monetary policy announcements ([35]), and political events ([86]). We prove that recommendation revisions made by efficient analysts contributing to efficient price discovery tend to not only generate significant stock market reactions but also reduce implied volatility and jump risks in option markets. Our new measure of SNR allows us to identify a market phenomenon that has not been proven in the literature, and we expect it can be used for other related purposes.

The remainder of the paper is organized as follows. Section 2.2 explains the methodology we use to describe the SNR measure along with the background literature as well as develops testable hypotheses. Section 2.3 describes our sample and provides the summary statistics of main variables in our study. Section 2.4 presents our main finding of this paper. We demonstrate the robustness of our main finding in Section

[5] Among many others, see [108], [1], and [13, 14] for the cases with the presence of microstructure noise. Under the assumption of no-microstructure noise, several studies use realized variance as a consistent estimator of integrated variance of the true efficient price process. For example, see [15] and [9] among many others.

2.5 and conclude in Section 2.6.

2.2 Methodology

In the literature on analyst recommendations, most studies use abnormal returns or trading volume around their issuances to examine the effect of recommendation revisions. These measures may contain noise, making it difficult to draw proper inferences on the fundamental role of security analysts: to gather, analyze, and ultimately provide information about a firm's value. In this section, we explain our approach to distinguish the impact of a recommendation regarding the unobservable true value of a firm from noise. We first suggest how to control for noise when measuring variation of true efficient prices reflecting the true value of a firm. Then, based on our proposal, we list the testable hypotheses to examine the impact of efficient analysts' recommendations on both stock and option markets.

2.2.1 Signal-to-Noise Volatility Ratio (SNR)

A firm value is represented by the true efficient price of a security. These efficient prices are not observable. In price data, they are often contaminated by noise due to market frictions. Essentially, our observed prices are a combination of these two unobservable components. In order to determine whether an analyst contributes to the discovery of true firm value or whether he/she induces more noise to observed prices, we suggest gauging the impact of analyst recommendations on the efficient price discovery by investigating the relative variations of the two components at the time of an analyst recommendation release. Our approach leverages recent developments in the literature on realized variance that separately estimates the variances of true efficient prices and noise.

We denote M and $\tilde{p}_{i,t,k}$ as the number of intraday discrete observations and the k-th logarithmic transaction, or quote price, of stock i on day t, respectively. This observed price $\tilde{p}_{i,t,k}$ consists of a true equilibrium price component $p_{i,t,k}$ related to a firm value and a microstructure noise component $\epsilon_{i,t,k}$ as follows:

$$\tilde{p}_{i,t,k} = p_{i,t,k} + \epsilon_{i,t,k}, \quad k = 1, 2, \cdots, M. \tag{8}$$

When we observe a change in observed price $\tilde{p}_{i,t,k}$ around the time when a recommendation revision is issued, it is difficult to distinguish whether the change is from the change in the efficient price $p_{i,t,k}$ or from the change in the noise component $\epsilon_{i,t,k}$. Our approach to resolve this difficulty is based on the intuition that the variation of $p_{i,t,k}$ will be affected more (less) than the variation of $\epsilon_{i,t,k}$ by the issuance of a revision if the revision contains more (less) valuable information about the true value of a firm. The relative ratio of these two variations tells us whether an analyst contributes to stock price efficiency.

To estimate the variation of each component in observed price $\tilde{p}_{i,t,k}$, we make the standard assumption on the true efficient price process. In particular, we assume that the logarithmic efficient price $p_{i,\tau}$ for stock i follows the diffusion process as in

$$p_{i,\tau} = p_{i,0} + \int_0^\tau \phi_{i,s} ds + \int_0^\tau \sigma_{i,s} dW_s, \tag{9}$$

where $\phi_{i,\tau}$ is a drift,[6] $\sigma_{i,\tau}$ is a spot volatility, and W_τ is a standard Brownian motion of stock i. If this continuous process representing a firm value is observable at discrete times in the absence of a noise component, one can express the k-th intraday logarithmic return of stock i on day t as

$$r_{i,t,k} = p_{i,t,k} - p_{i,t,k-1}, \quad k = 1, 2, \cdots, M. \tag{10}$$

Following the literature on realized variance, we define realized variance as the sum of all the intraday squared returns:

$$RV_{i,t} = \sum_{k=1}^{M} r_{i,t,k}^2. \tag{11}$$

[6]For our analysis using high-frequency data, the drift term is negligible compared to the volatility process due to the order of magnitude. We include it in Equation (9) for a generality.

If we assume the absence of a noise component in data, the realized variance is a consistent estimator for the integrated variance of the true efficient price process as we increase the frequency of intraday data.[7] That is,

$$RV_{i,t} - \int_0^t \sigma_{i,s}^2 ds \to 0, \qquad \text{as} \quad M \to \infty. \tag{12}$$

However, we know that the presence of a noise component is not negligible in observed prices. Then, an observed return can be represented as a sum of a change in a true equilibrium price and a change in a noise component:

$$\begin{aligned} \tilde{r}_{i,t,k} &= (\tilde{p}_{i,t,k} - \tilde{p}_{i,t,k-1}) \\ &= (p_{i,t,k} - p_{i,t,k-1}) + (\epsilon_{i,t,k} - \epsilon_{i,t,k-1}) \\ &= r_{i,t,k} + \eta_{i,t,k}. \end{aligned} \tag{13}$$

The observable price and efficient price can significantly diverge in reality due to liquidity or other frictions in financial markets (for example, see [104]). Thus, in the presence of noise, Equation (12) does not hold. As shown in [13], under the assumption of Equation (9) for the efficient price process and the assumption that the noise component $\epsilon_{i,t,k}$ is independent and identically distributed with a mean of zero and has a bounded fourth moment, the realized variance estimator does not consistently estimate the variance of the true efficient prices. Rather, the following result holds:

$$\sum_{k=1}^{M} \tilde{r}_{i,t,k}^2 \to \infty, \qquad \text{as} \quad M \to \infty. \tag{14}$$

That is, the sum of squared intraday returns contaminated by noise results in the infinite accumulation of microstructure noise as we increase the sampling frequency, M. Based on this result, a simple technique is provided by [13] to identify the variance of the efficient price process and the variance of microstructure noise using high-frequency data recorded at different frequencies.

[7] See [9] and [15] among others.

The technique is twofold. First, the variance (as well as other higher moments) of the unobserved noise process can be consistently estimated using the highest frequency data:

$$\frac{1}{M}\sum_{k=1}^{M}\tilde{r}_{i,t,k}^{q} \rightarrow \mathbb{E}(\epsilon_{i,t}^{q}), \quad q = 2,3,4 \qquad \text{as} \quad M \rightarrow \infty^{8}. \tag{15}$$

Second, using the sampling frequency of high-frequency data that optimally balances the finite sample bias and the variance of the estimator, we can extract the information about the variance of the true efficient price process from the traditional realized variance estimator [9]. Thus, we are able to infer the variance of market microstructure noise using the highest-frequency data as well as the variance of the true efficient price process using the optimally sampled high-frequency data.

We denote that, in the presence of microstructure noise, $\tilde{r}_{i,t,j}^{(\text{highest})}$ and $\tilde{r}_{i,t,j}^{(\text{opt})}$ are the j-th intraday logarithmic returns sampled at the highest frequency and at the optimal sampling frequency of stock i on day t following [13], respectively. Using these two different frequencies of returns, we further denote the estimator of the noise component volatility and the estimator of the efficient price volatility of stock i on day t as follows.

$$V_{i,t}^{(\text{highest})} \equiv \left[\frac{1}{J}\sum_{j=1}^{J}\tilde{r}_{i,t,j}^{(\text{highest})2}\right]^{1/2} \tag{16}$$

$$V_{i,t}^{(\text{opt})} \equiv \left[\frac{1}{K}\sum_{k=1}^{K}\tilde{r}_{i,t,k}^{(\text{opt})2}\right]^{1/2}, \tag{17}$$

where J and K are the numbers of intraday returns at the highest frequency and at the optimal sampling frequency, respectively $(J > K)$.

Using the results above, we construct our measure to examine the contribution of analysts' recommendations to the efficient price discovery, the *signal-to-noise volatility*

[8] See Theorem 2 and Appendix of [14] for details.
[9] See Section 4 of [13] and [14] for details.

ratio (SNR), as follows:

$$SNR_{i,t} \equiv \frac{V_{i,t}^{(\mathrm{opt})}}{V_{i,t}^{(\mathrm{highest})}}. \tag{18}$$

This efficiency measure allows us to gauge the relative variation in true efficient prices compared to the variation in noise when an analyst issues a recommendation revision. It indicates how informative the revision is regarding the true efficient price after controlling for the variance of the noise component. If an analyst recommendation contains valuable information on the true efficient price, the release would affect the variation of the efficient price process, leading to a relative increase of $V_{i,t}^{(\mathrm{opt})}$, and hence, an increase in $SNR_{i,t}$. However, if an analyst recommendation does not provide any material information about the true value of a firm, but only induces other market frictions, such as illiquidity or herding behavior of investors, it would result in an increased variation of noise, leading to a relative increase of $V_{i,t}^{(\mathrm{highest})}$, and hence, a decrease in $SNR_{i,t}$.

The main difference between this measure and other existing approaches to assess the impact of analysts is that our measure focuses on the fundamental role of efficient analysts, which is the discovery of efficient stock prices, whereas other approaches use a secondary outcome of analysts, such as stock price reaction, abnormal returns, or trading volumes. Furthermore, it takes into account the presence of noise prevailing in the observed market data. Therefore, we expect this measure to be robust to noise and useful for any analysis regardless of presence or amount of noise in data.

2.2.2 Testable Hypotheses for Informative Revisions

We evaluate the impact of recommendation revisions on stock and option markets where efficient prices of underlying assets are potentially affected by the release of efficient analysts' recommendation revisions. Since our measure helps us to differentiate more efficient analysts' recommendations from less efficient recommendations, we specifically investigate the degrees of market reactions in both markets. We can first confirm existing evidence that analyst recommendation revisions contribute to the discovery of the efficient stock price by considering the response of SNR around analyst revisions. Then, we set up three hypotheses for our study as follows.

In the stock market, if an analyst recommendation revision indeed contains valuable information about the true firm value, we expect the recommendation revision to be more efficient and so generate a greater degree of stock market reaction. This leads to our first testable hypothesis.

Hypothesis 1 *An analyst recommendation revision release with higher signal-to-noise volatility ratio (SNR), containing more valuable information about the efficient price of a security, generates a stronger market reaction and produces greater (abnormal) returns with the expected direction.*

In testing this hypothesis, we use upgrade and downgrade revisions separately to determine if efficient upgrade revisions generate greater abnormal returns, while efficient downgrade revisions generate lower abnormal returns. This test is important because ultimately the directional information of recommendation revisions matters for investors.

We also investigate how option markets respond to efficient analysts' recommendations. [98] document that the trading volume in derivate markets relative to the trading volume in underlying securities increases around information events, such as earnings announcements. In addition, [85] and [75] show that informed traders participate in option markets to exploit their informational advantages. If an analyst recommendation revision is an important informative event, we expect the trading volume in the option market relative to the trading volume in the stock market to increase on revision release days. Using our SNR measure, we can specifically test whether more efficient analysts' recommendations increase the relative option market volume more. This leads to our second hypothesis.

Hypothesis 2 *An analyst recommendation revision with higher signal-to-noise volatility ratio (SNR), containing more valuable information about the efficient price of a security, generates stronger option market trading activities.*

Lastly, we aim to examine if analyst recommendation revisions deliver material information to resolve uncertainty about the prospect of a firm. If a revision contains more information about the efficient price, reflecting the true value of a firm, more uncertainty is expected to be resolved, and hence, uncertainty is expected to decrease to a greater extent, relative to a less efficient revision. In the option market literature, the implied volatility or the slope of implied volatility has been used to measure uncertainty about the prospective of a firm. This leads to our third hypothesis.

Hypothesis 3 *An analyst recommendation revision with higher signal-to-noise ratio (SNR), containing more valuable information about the efficient price of a security, resolves uncertainty about a firm better, relative to a revision with lower SNR.*

2.3 Data and Sample

The sample period of our analysis spans from January 2001 to September 2014. The main variable of interest, the *signal-to-noise volatility* ratio (SNR), requires the use of two different frequencies of high-frequency data. We obtain tick-by-tick quote and trade data from the Trade and Quote (TAQ) database. In the empirical analysis, we focus on stocks in the Dow Jones Industrial Average Index.[10] From the composite of the Dow Jones Industrial Average Index, we cover the NYSE-listed 25 stocks because it is ideal for our purpose to use all available quotes from a consolidated exchange for consistency.

We obtain stock recommendation data from the Thomson Financial's Institutional Brokers Estimate (I/B/E/S) U.S. detail file. Analyst recommendation release

[10]In our sample, we include the stocks in the Dow Jones Industrial Average Index as of September 2014.

dates are also obtained from the same file. The I/B/E/S standardizes analyst recommendations and converts them into numerical scores, where "1" is strong buy, "2" is buy,"3" is hold, "4" is underperform, and "5" is sell. We reverse them so that higher numerical values correspond to more favorable recommendations. Previous research documents that recommendation revisions are more informative than levels of recommendations (see, for example, [21] and [82] among many others), so we focus on recommendation revisions computed as the difference between the current rating and the previous rating. Data on implied volatility and trading volume of individual options are obtained from OptionMetrics. We also collect data on market capitalization, trading volume, daily bid-ask spreads, and other stock price related data from the Center for Research in Security Prices (CRSP) database. Accounting data, such as the book value of firms, are obtained from the COMPUSTAT database.

To construct the *signal-to-noise volatility* (SNR) ratio, we first apply a common filter to remove quotes whose associated price changes or spreads were larger than 10% following [13]. Using the filtered quote data, we aggregate quotes to the second by averaging midprices of all quotes arriving within each second. Then, we use these one-second frequency quotes to estimate the variance of the microstructure noise process, $V_{i,t}^{(1sec)}$, using Equation (16).

[13] use the conditional mean-squared error (MSE) to estimate the variance of the true efficient price of a security. To evaluate the optimal sampling frequency for each security, they balance the trade-off between bias and the variance of an estimator. In our analysis, instead of estimating optimal sampling frequencies for each security, we construct five-minute returns as differences between logarithmic prices sampled every five-minutes. [13] analyze stocks in the S&P 100 index in the month of February 2002. The average, minimum, and maximum of optimal sampling frequencies of the stocks in their sample are around 4.0, 0.4, and 13.6 minutes, respectively. The average optimal sampling frequencies of securities that overlap in our sample is 3.9 minutes.

We use a five-minute frequency as an optimal frequency for our analysis to estimate a realized volatility of the true efficient price process, $V_{i,t}^{(5min)}$, using Equation (17).[11] We construct our main measure, the SNR, at a daily basis as a ratio of $V_{i,t}^{(5min)}$ to $V_{i,t}^{(1sec)}$ for all stocks in our sample: $SNR_{i,t} = \frac{V_{i,t}^{(5min)}}{V_{i,t}^{(1sec)}}$.

Table 15 presents descriptive statistics of variables used in this paper by each year and by each security in Panel A and Panel B, respectively. We report averages of daily closing prices, daily returns, volatility estimates for the true efficient prices ($V^{(5min)}$), volatility estimates for the noise ($V^{(1sec)}$), SNRs, daily stock trading volumes, daily option trading volumes, O/S ratios, market capitalizations, book-to-market ratios, and [5] illiquidity measures in each column. The time-series averages of the true efficient price volatility and the noise process volatility are 21.38% and 0.12%, and the average of the SNRs is 176.2. In Panel A of Table 15, we observe a cross-sectional variation of the SNRs, where Walt Disney (Symbol: DIS) shows the lowest ratio of 166.9 and UnitedHealth Group (Symbol: UNH) shows the highest ratio of 186.3. With the face values of these numbers, assuming the independence between the true efficient price process and the noise component, 0.0032% ($= 1/(176.2)^2$) of the variance of observed prices is due to the variance of the noise component when we sample 5-minute returns.

In Table 16, we report descriptive statistics of our main variable of interest, the SNR. From this point, we take the log of the SNR, as the ratio is strongly positively skewed. For our regression analyses, we use the log of the SNR ($\log(SNR)$). Panel A of Table 16 provides the summary statistics of $\log(SNR)$ for each firm in our sample. The mean and the median of $\log(SNR)$ are 5.153 and 5.159, respectively. In Panel

[11] Our choice of a five-minute sampling frequency for the true efficient price variance estimation can be considered conservative. Following [13], we randomly select a subset of stocks studied in this paper and estimate the optimal sampling frequencies for those stocks each month. The average, minimum, and maximum optimal sampling frequencies are 3.6, 3.03 (Caterpillar Inc. Symbol: CAT) and 4.02 (American Express Company. Symbol: AXP) minutes, respectively. Even though optimal sampling frequencies are relatively high in recent periods, the 5-minute sampling frequency is a conservative choice to prevent the variance estimator from contamination due to the noise component.

B of Table 16, we report the correlation coefficients matrix of variables, including $\log(SNR)$. $\log(SNR)$ shows positive correlations with stock trading volume, option trading volume, and bid-ask spread. The correlation coefficient between $\log(SNR)$ and daily returns is 0.0027, which is not statistically different from zero. This small correlation tells that the construction of our efficiency measure, $\log(SNR)$, is not mechanically related to the daily return of a stock.[12]

2.4 Empirical Findings

In this section, we first investigate how our measure, the SNRs, behaves around the timing of recommendation revision issuances. Once we formally examine the relationship between our measure of efficiency contribution and the issuances of recommendation revisions, we then further investigate efficient analysts' impacts on stock markets relative to noisy analysts. Next, we examine the impact of revisions contributing to stock price efficiency on option market activity and uncertainty resolution in option markets.

2.4.1 Recommendation Revisions and Stock Price Efficiency

We examine how the SNR is associated with the issuance of recommendation revisions. On average, we would expect the SNRs to be strongly associated with the issuance of analyst recommendation revisions if analysts work as important contributors to stock price efficiency. To check this, we merge the SNRs with the recommendation revisions data. Before formally testing the association using regression analysis, we first examine how recommendation revisions are related to the SNRs, raw daily returns, abnormal daily returns, and O/S ratios.

[12]We also checked the correlation of $\log(SNR)$ with an absolute value of a daily stock return, and the correlation coefficient is 13.36%. If we restrict our sample with large price changes, for example observations with $\pm 5\%$ or $\pm 10\%$ return in a day, then the correlation coefficients are not statistically different from zero, which again confirms no mechanical relation between the construction of $\log(SNR)$ and daily returns.

Table 17 reports the averages of the SNRs, raw daily returns, abnormal daily returns from [55], and O/S ratios. Panel A reports the results for daily observations that are not accompanied by recommendation revisions. Panel B and Panel C report the results for those that are accompanied by downgrade and upgrade revisions, respectively. Comparing the SNRs in Panel A with those in other panels, we observe increases in the SNRs when the observations are accompanied by recommendation revisions. The daily returns and the abnormal returns in Panel B and Panel C are in line with prior research that shows recommendation revisions have a significant impact on stock returns.[13] We also find evidence that information inflow due to revisions is exploited in the option market by observing an increase in O/S ratio around recommendation revision issuances. The finding on the O/S ratio implies increased activity of informed traders in the option market around the revision issuances, consistent with [98]. We investigate this finding in more detail in Subsection 2.4.3.

Figure 3 illustrates the dynamics of the SNR around recommendation revisions. The variation of the SNRs suggests that, on average, recommendation revisions contain valuable information about the true efficient price of a security as we observe the spikes of the SNRs on the days of revision issuances. To further confirm this finding, we test the following regression model at a firm-day observation level:

$$\log(SNR_{i,t}) = \beta \times d\_revision_{i,t} + \gamma' X_{i,t} + \theta_i + \delta_t + \varepsilon_{i,t}, \tag{19}$$

where the dependent variable, $\log(SNR_{i,t})$, is the log of the SNR for firm i on day t, and $d\_revision_{i,t}$ includes either $d\_up_{i,t}$ or $d\_down_{i,t}$, which is equal to one if an observation is accompanied by a downgrade or an upgrade revision, respectively. Alternatively, we include both $d\_up_{i,t}$ and $d\_down_{i,t}$ to estimate coefficients simultaneously. Because the contribution of a revision to stock price efficiency might be

[13]For example, [107] shows that recommendations of security analysts generate significant abnormal returns around the new buy and sell recommendations.

associated with firm characteristics, we include various control variables in the regression. $X_{i,t}$ represents a vector of control variables for firm i on day t, including the size ($\log(ME)$), the book-to-market ratio ($\log(BM)$), the trading volume ($\log(vol)$) and the [5] illiquidity measure ($Amihud$). Detailed definitions of variables are provided in Appendix B.1. We include firm fixed effects (θ_i) to control for any time-invariant unobservable heterogeneity across firms and day fixed effects (δ_t) to control for time-varying factors that might drive the changes of the SNRs for all firms.

Table 18 shows the results for the regression specification in Equation (19). The positive and statistically significant coefficients on d_down and d_up indicate that the SNR increases when recommendation revisions are released. In columns (3), (5), and (6) of Table 18, we find that the significance remains intact after controlling for firm characteristics and the illiquidity measure. The results suggest that the recommendation revisions of analysts in general contain valuable information on the true efficient price of a security.

2.4.2 Efficient Analysts' Recommendations and the Stock Market

Prior research has considered the value of analyst recommendations based on their impact on observed stock market returns[14], which could be possibly confounded with liquidity issues or herding behavior of stock market participants. Therefore, in this paper, we argue that the true value of a recommendation revision should be evaluated by considering the contribution of analysts to the efficient stock prices, which is a fundamental role of analysts in financial markets.

In this section, we test Hypothesis 1 as to whether recommendation revisions with more information regarding the true efficient price (i.e., higher SNRs) generate greater abnormal stock returns in the expected directions. We run the following regression

[14] Among many others, for example, see [107], [81], and [92].

specification at the firm-day observation level:

$$AR_{i,t} = \beta_1 \times d\_revision_{i,t} + \beta_2 \times d\_revision_{i,t} \times \log(SNR_{i,t}) + \theta_i + \gamma' X_{i,t} + \varepsilon_{i,t}, \quad (20)$$

where the dependent variable, $AR_{i,t}$, is an abnormal return of firm i on day t from the [55] three-factor model. $\log(SNR_{i,t})$ is the log of the SNR for a firm i on day t, and $d\_revision_{i,t}$ includes either $d\_up_{i,t}$ or $d\_down_{i,t}$, which is equal to one if an observation is accompanied by a downgrade or an upgrade revision, respectively. Alternatively, we include both $d\_up_{i,t}$ and $d\_down_{i,t}$ to estimate coefficients simultaneously. $X_{i,t}$ represents a vector of control variables for firm i on day t, including the size ($\log(ME)$), the book-to-market ratio ($\log(BM)$), the trading volume ($\log(vol)$) and the [5] illiquidity measure (*Amihud*). Detailed definitions of variables are provided in Appendix B.1. [66] suggest using a firm-fixed effect even when one uses a stock return as a dependent variable to control for unobserved heterogeneity. Following [66], we include firm fixed effects (θ_i) to control for any time-invariant unobservable heterogeneity across firms. The regression coefficient of interest β_2 would be expected to be negative for downgrade revisions and positive for upgrade revisions because we expect that downgrades and upgrades contributing to the improvement in price efficiency tend to generate greater negative and greater positive abnormal stock returns, respectively.

Consistent with our expectation, we find evidence that only revisions contributing to the discovery of the efficient stock prices generate abnormal returns in the expected directions. Table 19 provides the results for the regressions using Equation (20). The regression coefficients in columns (1) and (3) are negative and positive, respectively, and both are statistically significant. This result is consistent with the existing findings that downgrade and upgrade revisions generate negative and positive abnormal returns on average. Interestingly, when we include $\log(SNR)$ and interact it with the downgrade and upgrade dummies, it does not generate the expected stock market reaction. In column (5), we find that the coefficient on d_down is positive and significant while the coefficient estimate on $d\_down \times \log(SNR)$ is negative and significant. In column (6), only the interaction term of the upgrade revision dummy with the SNR shows positive significance.

To further tease out how different levels of the SNRs of revisions result in different responses in the stock market, we construct and analyze the subsamples separately: the downgrade revision sample and the upgrade revision sample. In addition, we create dummy variables, d_low, d_mid, and d_high, which are equal to one if an observation is in the bottom, middle, or top terciles of $\log(SNR)$, respectively. With these subsamples and the dummies for terciles of the SNRs, we test whether recommendation revisions with higher SNRs generate greater abnormal returns in the expected direction. The regression results are provided in Table 20. The coefficient estimates in columns (1), (2), (6), and (7) suggest that, as documented in the literature, downgrade revisions usually exert a greater impact on stock returns than upgrade revisions. In columns (3), (4), and (5), we find that only efficient analysts' downgrade revisions with high SNRs generate greater negative reactions in the stock market, whereas noisy analysts' downgrade revisions with low SNRs generate unexpected positive stock market reactions. These results demonstrate how our measure can be utilized to distinguish efficient analysts' downgrade revisions from noisy analysts' downgrade revisions.

2.4.3 Efficient Analysts' Recommendations and the Option Market

According to our finding in the previous subsection, efficient analysts' recommendations affect stock market returns in the anticipated directions. Our first question in this subsection is how informed traders would react around days when analysts release their recommendation revisions. If recommendations are believed to provide valuable information to infer the true efficient price of a security, these traders are expected to execute their transactions where the potential profit is maximized. That is, trading options on individual stocks would be much more appealing and preferred than trading underlying stocks. Since we can identify recommendation revisions contributing to the improvement in stock price efficiency using our SNR measure, we test how the issuances of those efficient analysts' revisions are related to the corresponding stocks' option market activities.

Another important benefit of considering option market responses is that it allows us to measure ex-ante market uncertainty. Implied volatility is often used to measure forward-looking volatility risk. The slope of implied volatility can also be used to measure extreme uncertainty going beyond regular volatility risk, which is referred to as jump or tail risk in the literature. These ex-ante uncertainty measures change over time as information flows. Our questions in this subsection are whether these forward-looking measures are affected by the efficient analysts' recommendations and whether those recommendations play a role in resolving relevant uncertainty. If a revision delivers important information about uncertain aspect of firm fundamentals, the release of a revision should help to resolve uncertainty, and hence, the subjective distribution for pricing related stock options is expected to exhibit lower volatility risk and a smaller magnitude of jump risk. Therefore, both implied volatility and its slope are expected to decrease when the revisions with higher SNRs are released. To the best of our knowledge, this is the first study which investigates the impact of analysts' recommendations on individual equity option market activities and uncertainty resolution.

2.4.3.1 Efficient Analysts' Recommendations and Option Market Activity

As informed traders are market participants who exploit informational advantages in their investment decisions, the revisions with highly valuable information about the efficient stock prices are likely to be utilized by them. In this subsection, we investigate the impact of efficient analysts' recommendation revisions on option market activity by testing Hypothesis 2.

We consider option market activity following [98], [85], [75], and many others on informed trading and option market participation. As aforementioned papers document that informed traders are more likely to actively participate in option markets, we expect that revisions with greater fundamental information regarding the true efficient price generate greater option trading volume relative to stock trading volume. To test the above, we run the following regression specification at a firm-day observation level:

$$\log(O/S\,ratio_{i,t}) = \beta_1 \times d\_revision_{i,t} + \beta_2 \times d\_revision_{i,t} \times \log(SNR_{i,t}) + \theta_i + \gamma' X_{i,t} + \varepsilon_{i,t}, \tag{21}$$

where the dependent variable, $\log(O/S\ ratio_{i,t})$, is the log of the ratio of option trading volume to stock trading volume of firm i on day t. As documented in [98], we use the O/S ratio to capture investor activity in the option market relative to the stock market. $\log(SNR_{i,t})$ is the log of the SNR for firm i on day t, and $d\_revision_{i,t}$ includes either $d\_up_{i,t}$ or $d\_down_{i,t}$, which is equal to one if an observation is accompanied by a downgrade or an upgrade revision, respectively. Alternatively, we include both $d\_up_{i,t}$ and $d\_down_{i,t}$ to estimate coefficients simultaneously.

In this subsection on option market analysis, a set of control variables ($X_{i,t}$) include stock market variables as well as option market variables to reduce the concern on omitted variables. As stock market control variables, we include the size ($\log(ME)$), the book-to-market ratio ($\log(BM)$), the bid-ask spread of a stock (Stock Spread(%)), and the [5] illiquidity measure (Amihud). As [98] examine economically plausible determinants of the O/S ratio, we also include those determinants as control variables: average option bid-ask spread (Option Spread(%)), average delta of options (Delta, with put deltas being reversed in sign), analyst forecast dispersion (Analyst Disp), institutional ownership (IO ratio), and average implied volatility (avg Ivol). All variables are defined in detail in Appendix A. Again, we include firm fixed effects (θ_i)

to control for any time-invariant unobservable heterogeneity across firms.

We document an increase in informed trader activity in option markets only around the times when efficient analysts' recommendation revisions are released. Table 21 reports the results for the regression specification (21). The positive and significant coefficients on d_down and d_up indicate that downgrade and upgrade revisions increase option market activity relative to stock market activity. However, when we include an interaction term of the revision dummies (d_down and d_up) and $\log(SNR)$ in the regression specification (columns (3) and (6)), the coefficients on the revision dummies (d_down and d_up) become negative, while those on the interaction terms ($d\_down \times \log(SNR)$ and $d\_up \times \log(SNR)$) are positive and significant. These results show that, consistent with Hypothesis 2, recommendation revisions containing more valuable information regarding the true efficient price tend to increase option trading volume compared to stock trading volume, while noisy analysts' revisions do not.

2.4.3.2 Efficient Analysts' Recommendations and Uncertainty Resolution

We have established that recommendation revisions contributing to the improvement in stock price efficiency generate stronger reactions in both stock and option markets. Since efficient analysts' revisions help investors infer the true efficient price of a security, we are able to further extend our analysis of the impact of those revisions on uncertainty resolution regarding the future prospects of a firm. If a recommendation revision provides important information to deduce the true value of a firm, the newly released information would help resolve uncertainty about a firm's future prospect. In this subsection, we formally test Hypothesis 3.

The price of an option contract reflects the expectation of market participants on future stock price movements, allowing us to study the uncertainty resolution around a firm's outlook. Among many other variables in option market data, the implied volatility of an option price is a natural proxy for the uncertainty measure (see [48], for example). To test whether efficient analysts' revisions reduce the future uncertainty of a firm using implied volatility, we run the following regression at a firm-day observation level:

$$\log(ImpVol_{i,t}) = \beta_1 \times d\_revision_{i,t} + \beta_2 \times d\_revision_{i,t} \times \log(SNR_{i,t}) + \theta_i + \gamma' X_{i,t} + \varepsilon_{i,t}, \tag{22}$$

where the dependent variable, $\log(ImpVol_{i,t})$, is the log of implied volatility, which is measured as the average of implied volatilities of at-the-money (ATM) call and put options (with 30 days of time-to-maturity) of a firm i on day t obtained from OptionMetrics. $\log(SNR_{i,t})$ is the log of the SNR for firm i on day t, and $d\_revision_{i,t}$ includes either $d\_up_{i,t}$ or $d\_down_{i,t}$, which is equal to one if an observation is accompanied by a downgrade or an upgrade revision, respectively. Alternatively, we include both $d\_up_{i,t}$ and $d\_down_{i,t}$ to estimate coefficients simultaneously. The same control variables are included in $X_{i,t}$ as in the regression specification (21), except the average of implied volatility (avg Ivol).

We find evidence that downgrade revisions containing valuable information on the true efficient price resolve uncertainty about a firm's future prospect. Table 22 provides the results for the regression analysis using Equation (22). In columns (2) and (5) of Table 22, we find that recommendation revisions decrease the average of implied volatilities of call and put options. However, when we include an interaction term of revision dummies (d_down and d_up) and $\log(SNR)$ in the regression specification (columns (3) and (6)), we find that only highly efficient analysts' downgrade revisions reduce implied volatility.[15]

[15] This result may appear difficult to reconcile with the leverage effect in which a decline in prices is associated with higher volatility due to the firm becoming more levered. We note here that higher volatility in leverage effect is realized volatility, while our measure of uncertainty in this analysis is implied volatility which is not realized but depends on investors subjective distribution of future stock returns. The subjective return distribution is based on the information up to the time of the issuance of a recommendation revision and does not necessarily determine realized volatility after

When there exists considerable uncertainty about a firm's future outlook, investors may anticipate high volatility and sudden jumps or discontinuities in stock prices in the near future. Thus, the risk (or likelihood) of jumps or discontinuous changes in stock prices is another signal indicating significant uncertainty in the firm's outlook. If a recommendation revision contains valuable information on the true efficient price of a stock, it would also reduce the likelihood of jumps or sudden discontinuities in future stock prices. That is, when an efficient analyst issues a recommendation revision, the jump risk in the price process would decrease due to uncertainty resolution from the issuance (Hypothesis 3).

In option markets, the prices of out-of-the-money (OTM) put options are sensitive to extreme movements in underlying security prices. For this reason, [86] use the slope of implied volatility to show that political events resolve the risk of jumps in stock prices. Thus, to test Hypothesis 3, we run the following regression specification:

$$Slope\_IVS_{i,t} = \beta_1 \times d\_revision_{i,t} + \beta_2 \times d\_revision_{i,t} \times \log(SNR_{i,t}) + \theta_i + \gamma' X_{i,t} + \varepsilon_{i,t}, \tag{23}$$

where the dependent variable, $Slope\_IVS_{i,t}$, is the slope of the implied volatility smile of options whose underlying asset is a stock of firm i on day t. The slope of the implied volatility smile is measured by running regressions of implied volatilities of OTM put options (with 30 days of time-to-maturity) on the deltas of corresponding options, following [86]. $\log(SNR_{i,t})$ is the log of the SNR for firm i on day t, and $d\_revision_{i,t}$ includes either $d\_up_{i,t}$ or $d\_down_{i,t}$, which is equal to one if an observation is accompanied by a downgrade or an upgrade revision, respectively. Alternatively, we include both $d\_up_{i,t}$ and $d\_down_{i,t}$ to estimate coefficients simultaneously. The same control variables are included in $X_{i,t}$ as in the regression specification (21).

The regression results are reported in Table 23. In column (3), similar to our

the revision. Therefore, we interpret the decrease in implied volatility as a resolution of investors' perceived uncertainty.

finding in Table 22, we find a significantly positive coefficient on *d_down* and a significantly negative coefficient on *d_down* $\times \log(SNR)$. On the one hand, the positive coefficient on *d_down* suggests that noisy analysts' recommendation revisions increase jump risk in the price process. On the other hand, as the increase of the SNR due to recommendation revisions implies that the revision provides valuable fundamental information on the true efficient price, and efficient analysts' recommendation revisions not only resolves the uncertainty of a firm's future prospects but also reduces the jump risk in related markets in the future.

In Table 24, we focus on observations with downgrade or upgrade revisions to examine a direct relationship between $\log(SNR)$ and option market variables. In Panel A and Panel B, we report the coefficient estimates for the regressions of different option market measures, log(*O/S ratio*), $\log(ImpVol)$, and *Slope_IVS* on observations accompanied by downgrade revisions and observations accompanied by upgrade revisions, respectively. The positively significant estimates in columns (1), (2), (7), and (8) indicate that both downgrade and upgrade revisions with greater contribution to stock price efficiency induce stronger activity in the option market relative to the stock market. Consistent with our finding in Tables 22 and 23, the results from columns (3) to (6) and from columns (9) to (12) suggest that only highly efficient analysts' downgrade revisions resolve the future uncertainty, as measured by the implied volatility, and decrease the jump risk in prices, as measured by the slope of implied volatility smile.

Overall, the evidence from stock markets suggests that highly efficient analysts' downgrade (upgrade) revisions generate greater negative (positive) abnormal returns because those revisions help to discover information on the true efficient price of a security. The evidence from option markets indicates that efficient analysts' downgrade revisions resolve future uncertainty about a firm and reduce the jump risk in the price process, whereas upgrade revisions do not.

2.5 Robustness Checks

2.5.1 Concurrent Earnings Announcements

A firm's earnings announcement delivers important information to market participants about a firm's performance and prospects on future cash flows. [98] propose the O/S ratio as a measure of informed trading in option market relative to stock market increases around the times of earnings announcements of a firm. In addition, [44] document empirical evidence that implied volatility increases before earnings announcements and decreases after the announcements. Given that analyst recommendations are often issued around the times of earnings announcements, it is possible that our finding on efficient analysts' recommendations is a manifestation of the finding on earnings announcements. To mitigate this concern, we exclude firm-day observations used in our previous analysis that are accompanied by actual earnings announcements. Using this filtered sample, we first investigate the relationship between downgrade or upgrade recommendation revisions and $\log(SNR)$.

In Panel A of Table 25, we provide the results for the filtered sample without any earnings announcements. The results are very similar to the previous results in Table 18, implying that the revisions, on average, contain valuable information about the true efficient price of a security. Panel B and Panel C of Table 25 show the results for the impact of efficient analysts' revisions on stock and option markets. In Panel B, similar to Table 20, we examine whether recommendation revisions contributing to stock price efficiency generate greater abnormal returns than noisy analysts' recommendation revisions. Consistent with our finding in Table 20, we find that more efficient analysts' revisions generate greater negative abnormal returns in the stock market. In Panel C, we also find similar evidence as in Table 24. Namely, efficient analysts' recommendation revisions increase trading activity in the option market relative to that in the stock market, resolve uncertainty about a firm, and reduce the jump risk in the stock price of a firm. Thus, overall, our findings on the impact of efficient analysts' recommendation revisions on stock returns, option market activity, and uncertainty resolution are not driven by the confounding effect of earnings announcements.

2.5.2 Subsample Period Analysis

Since the construction of our main measure, the SNR, largely depends on high-frequency data and the presence of microstructure friction, there may be a concern that any structural breaks or abrupt changes in market microstructure might drive the variation of the SNR. In this subsection, we aim to mitigate such concerns by conducting a subsample period analysis. We divide our full sample into two subsamples, one from 2001 to 2007 and the other from 2008 to 2014, and perform the same analyses on each subsample.

In Panel A of Table 26, we report the results for the relationship between the SNR and recommendation revisions. Our finding is robust in both subsample periods as we find the issue of downgrade or upgrade revisions is associated with an increase in the SNR. The subsample results for the efficient analysts' revisions and stock returns are provided in Panel B, and we confirm that our finding in Section 2.4 is robust. Interestingly, we find evidence that our result becomes stronger (in terms of coefficients and t-statistics) in later years (2008-2014) than in earlier years (2001-2007). This is due to the fact that many uninformed stock market participants are now able to obtain various information including analysts recommendations more easily in recent years, which is consistent with the findings from the subsample period analysis.

Panel C of Table 26 provides the subsample results on option market activity and uncertainty resolution. We find that our main finding only holds in the later subsample period (2008-2014). This finding makes sense because the depth and the liquidity of the option market have been more well-developed in recent years as we observe an increase in option trading volume over time in Table 15.

2.6 Conclusion

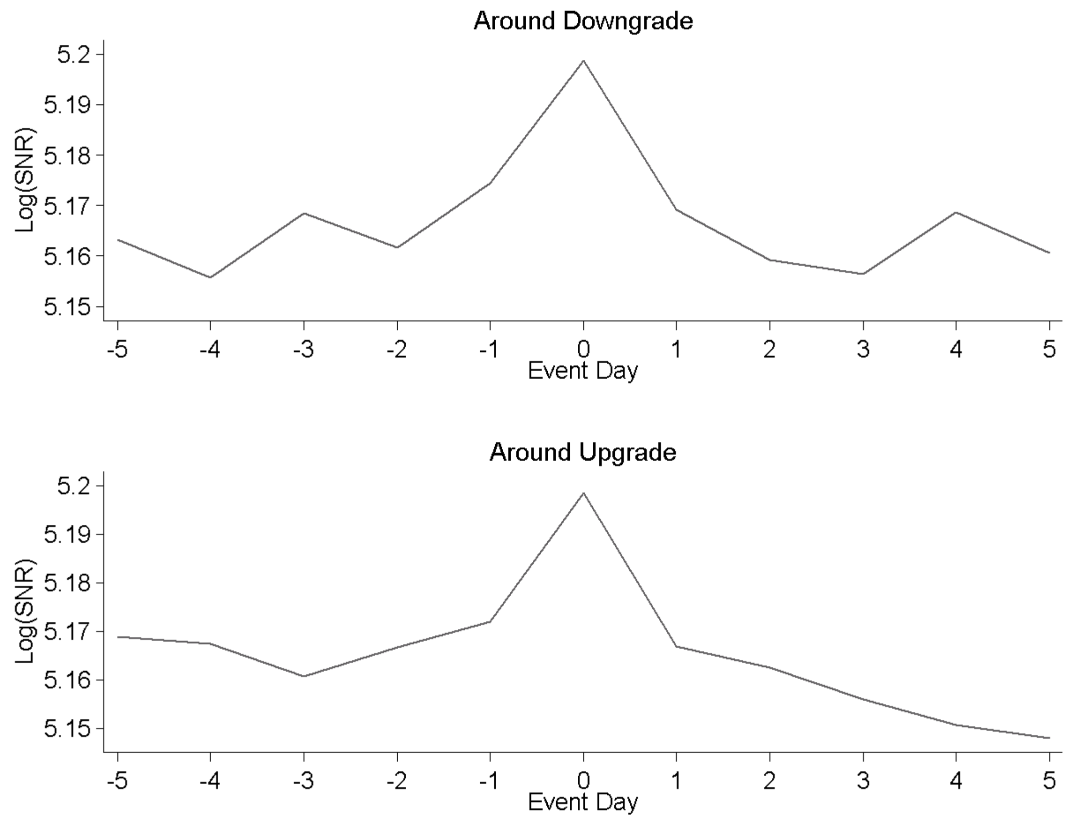
We investigate the fundamental role of security analysts in terms of their contribution to the stock price efficiency by considering their impact on the variation of the true efficient price after controlling for the presence of noise. Using two different frequencies of high-frequency data, we construct a measure, the *signal-to-noise volatility* ratio (SNR), which is defined as a ratio of the efficient price variation to the noise variation, and find that changes in stock price efficiency are strongly associated with the issuances of downgrade and upgrade revisions.

We distinguish recommendation revisions that contribute to the improvement in stock price efficiency (efficient analysts' recommendation revisions) from those that amplify the variance of the noise component in the observed stock prices (noisy analysts' recommendation revisions). We find that only highly efficient analysts' downgrade and upgrade revisions (i.e., high SNR revisions) produce significantly negative and positive abnormal returns, respectively. However, noisy analysts' downgrade revisions generate unexpected positive abnormal returns. Thus, we identify highly efficient analysts' revisions which generate significant stock returns as intended by the revisions.

We also find that efficient analysts' revisions are well exploited and reflected in option markets. First, we document that highly efficient analysts' revisions induce greater trading volumes in the option markets relative to the stock markets, which is consistent with the argument that informed traders actively participate in the option markets to exploit informational advantages. Second, we find evidence that highly efficient analysts' recommendation revisions help resolve future uncertainty of a firm and reduce the risk of jumps in stock prices as they provide important information to investors. Overall, the findings in this paper suggest that stock and option market participants rationally respond to the issuances of efficient analysts' recommendation revisions that reveal information about the true efficient price of a security.

Figure 3: Stock Price Efficiency Measure around Recommendation Revisions

The figures below display the measure of stock price efficiency around analyst recommendation revisions. The upper and lower plots show the dynamics of log(SNR) around downgrade and upgrade revisions, respectively.



As the first paper that focuses on the fundamental role of security analysts in discovering the true efficient price of a security, this paper sheds light on the use of high-frequency data to draw an inference on analyst recommendations regarding the true price discovery. Our finding further suggests that high-frequency data is a potent data source, which might enable us to solve other interesting economic questions.

Table 15: Descriptive Statistics

This table presents the descriptive statistics of the variables of interest in the paper. The sample contains 25 stocks in the Dow Jones Industrial Average Index from January 2001 to September 2014. All variables are at a daily basis for each firm. Panel A reports the annual averages of variables, and Panel B reports firm-level averages of variables. The two columns, Avg. price and Daily ret, report the average stock price and the average of daily returns. $V^{(1sec)}$ and $V^{(5min)}$ are the realized (annualized) volatilities using 1-second returns and 5-minute returns using Equation (16) and Equation (17), respectively. The *signal-to-noise volatility* ratio, SNR, is the ratio of $V^{(5min)}$ to $V^{(1sec)}$. Stock vol and Option vol show the average daily trading volume (in hundreds) of stocks and options, respectively. O/S ratio is the ratio of option trading volume to stock trading volume. The last three columns, ME, BM, and Amihud, are the averages of market capitalization, book-to-market ratio, and [5] illiquidity measure, respectively.

Panel A. By year

Year	Avg. price	Daily ret	$V^{(5min)}$	$V^{(1sec)}$	SNR	Stock vol	Option vol	O/S ratio	ME	BM	Amihud
2001	57.60	-0.0050%	0.3084	0.0017	183.6	60399.8	4006.7	0.057	112161.9	0.253	0.085
2002	53.74	-0.0450%	0.3137	0.0017	190.1	72539.8	4299.9	0.059	99542.9	0.323	0.115
2003	52.37	0.1222%	0.2124	0.0013	169.9	67754.6	4182.2	0.068	96332.2	0.333	0.084
2004	57.38	0.0601%	0.1641	0.0010	164.2	67084.8	3787.4	0.063	111980.6	0.286	0.056
2005	56.28	0.0166%	0.1611	0.0010	163.0	82841.7	5102.4	0.066	117016.8	0.324	0.057
2006	58.58	0.0757%	0.1625	0.0010	164.7	95527.8	7419.5	0.080	122151.8	0.343	0.056
2007	66.45	0.0401%	0.1936	0.0010	185.5	121853.0	8209.3	0.074	129103.1	0.341	0.093
2008	58.00	-0.0730%	0.3654	0.0021	177.3	179024.8	13890.0	0.076	112251.8	0.450	0.297
2009	50.20	0.1263%	0.2946	0.0017	177.7	178107.9	17251.9	0.096	92222.6	0.535	0.282
2010	60.61	0.0686%	0.1924	0.0010	186.0	149889.3	13669.3	0.105	103582.6	0.472	0.128
2011	68.22	0.0496%	0.2034	0.0011	181.3	133595.9	13308.1	0.118	115294.3	0.476	0.121
2012	74.54	0.0661%	0.1533	0.0009	174.4	105887.8	11900.1	0.131	124742.1	0.459	0.078
2013	86.15	0.1197%	0.1400	0.0008	176.3	92398.8	10655.6	0.140	143058.5	0.404	0.075
2014	95.90	0.0258%	0.1285	0.0008	172.3	83081.4	10440.3	0.149	153379.6	0.390	0.228
Total	64.00	0.0463%	0.2138	0.0012	176.2	106427.7	9151.6	0.092	116630.1	0.38	0.13

Panel B. By firm

Symbol	Avg. price	Daily ret	$V^{(5min)}$	$V^{(1sec)}$	SNR	Stock vol	Option vol	O/S ratio	ME	BM	Amihud
AXP	49.87	0.0515%	0.2531	0.0014	176.2	75146.9	5395.1	0.067	59680.7	0.25	0.21
BA	67.60	0.0458%	0.2363	0.0014	174.4	48657.6	5940.5	0.114	52206.1	0.31	0.12
CAT	71.58	0.0724%	0.2481	0.0014	182.9	58007.6	12274.7	0.171	39718.4	0.29	0.13
CVX	85.25	0.0561%	0.1977	0.0011	175.1	78024.8	6311.1	0.076	155812.4	0.54	0.08
DD	45.27	0.0412%	0.2264	0.0013	171.2	51997.9	2915.4	0.052	43051.6	0.37	0.16
DIS	34.79	0.0565%	0.2366	0.0014	166.9	96951.0	4068.8	0.044	66311.7	0.54	0.20
GE	27.82	0.0144%	0.2258	0.0013	177.1	381022.3	25259.2	0.056	284570.6	0.46	0.22
GS	131.26	0.0488%	0.2570	0.0014	178.4	68349.3	22321.1	0.292	60765.0	0.64	0.07
HD	41.60	0.0470%	0.2420	0.0014	180.8	117843.1	7449.5	0.065	76699.1	0.29	0.18
IBM	123.16	0.0411%	0.1873	0.0011	170.5	66857.6	16946.2	0.256	167373.4	0.19	0.05
JNJ	65.87	0.0374%	0.1572	0.0009	179.9	99672.1	4982.4	0.052	184463.9	0.23	0.06
JPM	40.65	0.0562%	0.2716	0.0015	177.5	233289.4	27619.1	0.100	135148.0	0.79	0.18
KO	49.79	0.0277%	0.1684	0.0010	174.0	88899.0	5334.3	0.055	130491.7	0.17	0.09
MCD	54.80	0.0500%	0.1966	0.0012	173.8	66389.3	6005.2	0.081	60468.5	0.25	0.11
MMM	93.34	0.0447%	0.1876	0.0011	173.4	33432.7	3082.6	0.098	59038.1	0.20	0.07
NKE	70.24	0.0757%	0.2241	0.0012	185.2	26179.2	2673.2	0.086	24543.9	0.34	0.11
PFE	26.46	0.0135%	0.2123	0.0012	174.6	329499.9	12655.3	0.037	187085.5	0.44	0.18
PG	69.66	0.0390%	0.1611	0.0009	174.8	87470.4	6949.3	0.067	166463.9	0.29	0.07
TRV	60.16	0.0618%	0.2214	0.0013	178.9	37093.8	547.1	0.016	27818.2	0.94	0.17
UNH	56.34	0.0733%	0.2513	0.0014	186.3	63512.5	4408.8	0.068	48297.1	0.35	0.17
UTX	75.66	0.0513%	0.2046	0.0012	167.2	38587.2	2302.5	0.056	60004.9	0.27	0.09
V	123.90	0.1306%	0.2054	0.0012	171.7	52160.0	10657.3	0.213	62689.4	0.51	0.09
VZ	38.82	0.0336%	0.2099	0.0012	173.4	125127.5	7940.3	0.051	111995.5	0.70	0.13
WMT	56.10	0.0264%	0.1862	0.0010	183.4	122475.8	10329.6	0.079	218598.9	0.26	0.10
XOM	68.35	0.0438%	0.1935	0.0011	178.6	184616.5	15587.0	0.082	351697.3	0.39	0.08

Table 16: Description of Signal-to-Noise Volatility Ratio

This table presents the descriptive statistics of the *signal-to-noise volatility* ratio (SNR), which is the main variable of interest in the paper. In this table, the log of SNR is computed as the log ratio of $V^{(5min)}$ to $V^{(1sec)}$. The sample contains 25 stocks in the Dow Jones Industrial Average Index from January 2001 to September 2014. In Panel A, we report the summary statistics of the log(SNR) for each firm. The mean, median, and standard deviation are reported in the second, third, and fourth columns, respectively. The minimum, maximum, 10th percentile, and 90th percentile are reported in the last four columns. In Panel B, the correlation matrix of the log(SNR) along with daily returns (Daily ret), stock trading volume (Stock vol), option trading volume (Option vol), firm size (log(ME)), book-to-market ratio (log(BM)), and bid-ask spread (Stock Spread(%)), is provided. * indicates statistical significance at the 10% level.

Panel A. Summary Statistics for Each Firm

Symbol	Mean	Median	SD	Min	Max	p10	p90
AXP	5.155	5.163	0.191	3.366	5.830	4.907	5.389
BA	5.144	5.146	0.183	3.705	6.005	4.914	5.371
CAT	5.195	5.194	0.165	3.999	5.985	4.986	5.399
CVX	5.148	5.152	0.193	4.065	6.226	4.914	5.387
DD	5.125	5.124	0.192	4.154	6.475	4.886	5.356
DIS	5.095	5.111	0.210	4.238	6.069	4.821	5.346
GE	5.162	5.164	0.167	3.684	6.016	4.956	5.373
GS	5.167	5.174	0.182	3.399	5.973	4.955	5.382
HD	5.179	5.186	0.193	3.821	5.943	4.943	5.414
IBM	5.120	5.125	0.199	3.818	6.214	4.870	5.365
JNJ	5.170	5.165	0.194	3.846	6.394	4.926	5.412
JPM	5.162	5.175	0.184	3.856	5.733	4.923	5.387
KO	5.139	5.141	0.192	4.172	5.931	4.897	5.374
MCD	5.137	5.138	0.195	3.744	6.212	4.905	5.370
MMM	5.135	5.157	0.212	3.307	6.002	4.858	5.380
NKE	5.203	5.205	0.188	3.789	6.390	4.972	5.424
PFE	5.145	5.148	0.185	4.279	6.015	4.910	5.374
PG	5.146	5.146	0.187	4.056	6.527	4.923	5.371
TRV	5.163	5.157	0.198	4.476	6.795	4.925	5.399
UNH	5.206	5.207	0.205	3.904	6.884	4.958	5.447
UTX	5.100	5.110	0.200	4.140	5.892	4.854	5.343
V	5.120	5.128	0.221	4.492	5.961	4.825	5.401
VZ	5.133	5.139	0.211	3.481	6.497	4.880	5.385
WMT	5.191	5.194	0.200	3.945	6.085	4.950	5.431
XOM	5.172	5.172	0.162	4.109	5.858	4.972	5.379
Total	5.153	5.159	0.194	3.307	6.884	4.912	5.387

Panel B. Correlation Matrix

	log(SNR)	Daily ret	Stock vol	Option vol	log(ME)	log(BM)
Daily ret	0.0027					
Stock vol	0.1125*	-0.0092*				
Option vol	0.0945*	0.0153*	0.4312*			
log(ME)	-0.005	0.0044	0.4089*	0.1789*		
log(BM)	-0.0044	-0.0126*	0.3212*	0.1676*	-0.1275*	
Stock Spread(%)	0.0584*	-0.0106*	-0.0755*	-0.0701*	-0.1138*	-0.0945*

Table 17: Impact of Recommendation Revisions on Signal-to-Noise Volatility Ratio, Daily Return, and O/S Ratio

This table presents the impact of analyst recommendation revisions on the *signal-to-noise volatility* ratio ($\log(SNR)$), daily raw and abnormal returns (Daily ret and Ab ret), and option-to-stock trading volume ratio (O/S ratio) for each firm in the sample. Panel A reports the averages of $\log(SNR)$, daily return, abnormal return from the [55] three-factor model, and O/S ratio when there is no analyst recommendation revision. Panel B (Panel C) provides the same variables (including the numbers of downgrade (upgrade) revisions under the column Obs) when downgrade (upgrade) revisions of analyst recommendations are accompanied. The sample period spans from January 2001 to September 2014.

	Panel A. No Revision				**Panel B. Downgrade Revision**					**Panel C. Upgrade Revision**				
Symbol	$\log(SNR)$	Daily ret	Ab ret	O/S ratio	$\log(SNR)$	Daily ret	Ab ret	O/S ratio	Obs	$\log(SNR)$	Daily ret	Ab ret	O/S ratio	Obs
AXP	5.15	0.03%	-0.02%	0.07	5.22	-0.71%	-0.83%	0.07	44	5.28	2.06%	1.08%	0.07	55
BA	5.14	0.05%	-0.01%	0.11	5.16	-0.81%	-0.93%	0.12	61	5.18	0.81%	0.80%	0.09	63
CAT	5.19	0.06%	0.00%	0.17	5.20	-0.37%	-0.60%	0.19	59	5.23	1.11%	0.81%	0.19	56
CVX	5.14	0.04%	-0.01%	0.08	5.21	-0.02%	-0.17%	0.07	60	5.22	1.02%	0.58%	0.07	55
DD	5.12	0.03%	0.00%	0.05	5.18	-0.89%	-0.45%	0.05	83	5.16	1.41%	0.75%	0.05	82
DIS	5.09	0.05%	-0.02%	0.04	5.19	-0.90%	-0.77%	0.05	60	5.18	1.39%	0.86%	0.06	67
GE	5.16	0.01%	-0.01%	0.06	5.22	-0.92%	-0.55%	0.05	40	5.22	1.27%	0.97%	0.06	30
GS	5.17	0.05%	-0.01%	0.29	5.23	-0.90%	-0.65%	0.30	70	5.18	1.06%	0.55%	0.30	68
HD	5.18	0.05%	0.00%	0.07	5.22	-1.38%	-0.78%	0.05	53	5.19	1.12%	0.93%	0.07	57
IBM	5.12	0.05%	0.00%	0.26	5.15	-1.15%	-0.71%	0.32	53	5.14	1.11%	0.62%	0.26	42
JNJ	5.17	0.04%	-0.01%	0.05	5.27	-0.82%	-0.63%	0.06	55	5.26	0.62%	0.50%	0.08	44
JPM	5.16	0.06%	-0.02%	0.10	5.23	-1.24%	-0.81%	0.11	44	5.17	0.77%	0.33%	0.10	51
KO	5.14	0.03%	0.00%	0.06	5.18	-1.04%	-0.85%	0.07	36	5.19	1.01%	0.67%	0.06	36
MCD	5.14	0.05%	0.00%	0.08	5.21	-0.57%	-0.61%	0.09	52	5.17	0.88%	0.84%	0.08	51
MMM	5.13	0.06%	0.01%	0.10	5.21	-1.37%	-1.04%	0.10	60	5.16	0.81%	0.79%	0.10	51
NKE	5.20	0.07%	-0.02%	0.08	5.25	-1.30%	-0.98%	0.11	35	5.23	1.22%	0.93%	0.14	46
PFE	5.15	0.02%	0.00%	0.04	5.11	-0.85%	-0.51%	0.04	53	5.11	0.41%	0.57%	0.04	47
PG	5.14	0.03%	-0.01%	0.07	5.21	-0.06%	-0.30%	0.07	47	5.23	0.57%	0.55%	0.06	43
TRV	5.16	0.07%	-0.02%	0.02	5.20	-0.79%	-0.84%	0.02	23	5.28	0.69%	0.61%	0.01	15
UNH	5.20	0.09%	-0.01%	0.07	5.30	-1.32%	-1.25%	0.07	48	5.30	0.40%	0.56%	0.08	43
UTX	5.10	0.05%	0.00%	0.06	5.12	-0.87%	-0.71%	0.05	25	5.15	0.86%	0.67%	0.05	28
V	5.12	0.15%	0.00%	0.21	5.16	-1.25%	-1.25%	0.21	23	5.24	0.22%	0.42%	0.23	15
VZ	5.13	0.05%	0.00%	0.05	5.18	-1.09%	-0.78%	0.05	81	5.22	0.60%	0.67%	0.05	82
WMT	5.19	0.04%	0.01%	0.08	5.25	-0.90%	-0.86%	0.07	79	5.19	0.56%	0.37%	0.08	63
XOM	5.17	0.03%	-0.01%	0.08	5.15	0.09%	-0.10%	0.09	74	5.20	0.99%	0.57%	0.09	62
Total	5.15	0.05%	-0.01%	0.09	5.20	-0.86%	-0.72%	0.10	1318	5.20	0.92%	0.68%	0.10	1252

Table 18: Analyst Recommendation Revisions and Stock Price Efficiency

In this table, we investigate how our measure of stock price efficiency is related to the issuances of analyst recommendation revisions. The table reports the regression result of the efficiency measure on dummy variables of upgrade and downgrade revisions with other control variables. The dependent variable, $\log(SNR)$, is the log of the *signal-to-noise volatility* ratio, which measures stock price efficiency. The dummy variables, *d_down* and *d_up*, are indicators equal to one if an observation is accompanied by a downgrade revision and an upgrade revision, respectively. $\log(ME)$, $\log(BM)$, $\log(vol)$, and Amihud are the market capitalization, book-to-market ratio, trading volume, and illiquidity measure of [5], respectively. All specifications in the table below include the firm- and the day-fixed effects, and the t-statistics are clustered at a daily level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Dependent Variable: $\log(SNR)$					
	(1)	(2)	(3)	(4)	(5)	(6)
d_down	0.0428***	0.0436***	0.0191***			0.0196***
	(8.08)	(8.24)	(3.73)			(3.84)
d_up		0.0412***		0.0403***	0.0189***	0.0195***
		(7.54)		(7.37)	(3.56)	(3.67)
$\log(ME)$			0.00996***		0.00991***	0.0101***
			(3.26)		(3.24)	(3.30)
$\log(BM)$			-0.0261***		-0.0262***	-0.0260***
			(-11.87)		(-11.88)	(-11.82)
$\log(vol)$			0.0858***		0.0859***	0.0853***
			(43.33)		(43.32)	(42.99)
Amihud			0.131***		0.131***	0.131***
			(13.89)		(13.90)	(13.90)
Firm-Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Day-Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	82627	82627	82627	82627	82627	82627
R-squared	0.218	0.219	0.251	0.218	0.251	0.251

Table 19: Efficient Analysts' Recommendation Revisions and Stock Market Reaction

The table investigates the impact of efficient analysts' recommendation revisions on the stock market reaction. The table reports the regression result of abnormal stock returns on dummy variables of upgrade and downgrade revisions interacting with the *signal-to-noise volatility ratio* (SNR). The dependent variable, AR(0,0), is an abnormal return from the Fama and French (1993) three-factor model. The dummy variables, *d_down* and *d_up*, are indicators equal to one if an observation is accompanied by a downgrade revision and an upgrade revision, respectively. $\log(SNR)$ is the log of the *signal-to-noise volatility* ratio, which measures stock price efficiency. $\log(ME)$, $\log(BM)$, $\log(vol)$, and Amihud are the market capitalization, book-to-market ratio, trading volume, and illiquidity measure of [5], respectively. All specifications below include the firm-fixed effect, and the standard errors are clustered at a daily level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Dependent Variable: AR(0, 0)					
	(1)	(2)	(3)	(4)	(5)	(6)
d_down	-0.00692***	0.0570***			0.0588***	
	(-15.95)	(5.22)			(5.34)	
d_down× $\log(SNR)$		-0.0123***			-0.0127***	
		(-5.82)			(-5.94)	
d_up			0.00707***	-0.0160		-0.0152
			(15.90)	(-1.44)		(-1.37)
d_up× $\log(SNR)$				0.00443**		0.00428**
				(2.05)		(1.98)
$\log(ME)$					-0.0000329	0.000125
					(-0.21)	(0.79)
$\log(BM)$					-0.000527***	-0.000381***
					(-3.67)	(-2.65)
$\log(vol)$					0.000520***	0.000213*
					(4.38)	(1.80)
Amihud					0.000133	0.000261
					(0.24)	(0.47)
Firm-Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	82576	82576	82576	82576	82576	82576
R-squared	0.006	0.007	0.006	0.006	0.007	0.006

Table 20: Subsample Analysis: Efficient Analysts' Recommendation Revisions and Stock Market Reaction

This table investigates the non-linear impact of efficient analysts' recommendation revisions (downgrade and upgrade, separately) on the stock market reaction. The table reports the regression result of abnormal stock returns on the *signal-to-noise volatility* ratio (SNR). The dependent variable, AR(0,0), is an abnormal return from the Fama and French (1993) three-factor model. Panel A and Panel B show the results for the subsample with downgrade revisions and the subsample with upgrade revisions, respectively. $\log(SNR)$ is the log of the *signal-to-noise volatility* ratio, which measures stock price efficiency. To examine the level of efficiency contribution of revisions and its impact on the abnormal returns, we create the dummy variables, *d_low*, *d_mid*, and *d_high*, which are equal to one if an observation is in bottom, middle, or top terciles of $\log(SNR)$, respectively, and interact them with $\log(SNR)$. $\log(ME)$, $\log(BM)$, $\log(vol)$, and Amihud are the market capitalization, book-to-market ratio, trading volume, and illiquidity measure of [5], respectively. All specifications below include the firm-fixed effect, and the standard errors are clustered at a daily level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Dependent Variable: AR(0, 0)									
	Panel A. Downgrade Revisions					Panel B. Upgrade Revisions				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$\log(SNR)$	-0.0144***	-0.00671**				0.00523**	0.00126			
	(-5.23)	(-2.50)				(2.04)	(0.49)			
$\log(SNR)$ × d_low			0.000464**					0.0000416		
			(2.30)					(0.21)		
$\log(SNR)$ × d_mid				0.0000495					-0.000141	
				(0.26)					(-0.70)	
$\log(SNR)$ × d_high					-0.000488**					0.000107
					(-2.46)					(0.52)
$\log(ME)$		0.00400**	0.00404**	0.00402**	0.00410**		-0.00343	-0.00340	-0.00338	-0.00343
		(2.09)	(2.11)	(2.09)	(2.14)		(-1.61)	(-1.59)	(-1.58)	(-1.61)
$\log(BM)$		0.00354**	0.00372**	0.00387**	0.00364**		-0.00412**	-0.00420**	-0.00413**	-0.00410**
		(2.17)	(2.30)	(2.40)	(2.26)		(-2.24)	(-2.29)	(-2.25)	(-2.23)
$\log(vol)$		-0.0115***	-0.0118***	-0.0121***	-0.0116***		0.00742***	0.00753***	0.00750***	0.00744***
		(-9.52)	(-9.90)	(-10.30)	(-9.70)		(5.15)	(5.26)	(5.28)	(5.15)
Amihud		0.00723	0.00738	0.00742	0.00700		-0.00118	-0.00114	-0.000995	-0.00113
		(1.49)	(1.53)	(1.52)	(1.44)		(-0.26)	(-0.25)	(-0.22)	(-0.25)
Firm-Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1318	1318	1318	1318	1318	1252	1252	1252	1252	1252
R-squared	0.056	0.163	0.162	0.158	0.162	0.020	0.070	0.070	0.070	0.070

Table 21: Efficient Analysts' Recommendation Revisions and Option Market Activity

This table investigates the impact of efficient analysts' recommendation revisions on option market activity. The table reports the regression results of option market trading activity on dummy variables of upgrade and downgrade revisions interacting with the *signal-to-noise volatility* ratio (SNR). The dependent variable, $\log(O/S\,ratio)$, is the log of the ratio of option trading volume to stock trading volume to capture investor activity in the option market relative to the stock market. The dummy variables, *d_down* and *d_up*, are indicators equal to one if an observation is accompanied by a downgrade revision and an upgrade revision, respectively. $\log(SNR)$ is the log of the *signal-to-noise volatility* ratio, which measures stock price efficiency. $\log(ME)$, $\log(BM)$, *Amihud*, and Stock Spread(%) are the market capitalization, book-to-market ratio, illiquidity measure of [5], and percentage of bid-ask spread of underlying stock, respectively. Option related controls are the percentage of option bid-ask spread (Option Spread(%)), average of option deltas (Delta, with put deltas being reversed in sign), analyst forecast dispersion (Analyst Disp), institutional ownership (IO ratio), and average of implied volatility (avg Ivol). All specifications below include the firm- and the day-fixed effects, and the standard errors are clustered at a daily level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Dependent Variable: $\log(O/S\,ratio)$					
	(1)	(2)	(3)	(4)	(5)	(6)
d_down	0.0751***	0.0668***	-1.607***			
	(3.51)	(3.31)	(-2.84)			
d_down×$\log(SNR)$			0.322***			
			(2.98)			
d_up				0.118***	0.106***	-1.246**
				(4.96)	(4.74)	(-2.20)
d_up×$\log(SNR)$						0.260**
						(2.39)
$\log(ME)$		0.469***	0.469***		0.469***	0.469***
		(31.97)	(31.95)		(31.95)	(31.93)
$\log(BM)$		-0.352***	-0.352***		-0.352***	-0.352***
		(-33.99)	(-34.00)		(-34.02)	(-34.02)
Amihud		-0.339***	-0.341***		-0.338***	-0.339***
		(-9.67)	(-9.69)		(-9.62)	(-9.64)
Stock Spread(%)		-0.209***	-0.209***		-0.208***	-0.208***
		(-11.23)	(-11.23)		(-11.22)	(-11.22)
Option Spread(%)		-0.268***	-0.268***		-0.269***	-0.268***
		(-4.29)	(-4.29)		(-4.30)	(-4.29)
Delta		0.292***	0.292***		0.289***	0.290***
		(3.08)	(3.08)		(3.04)	(3.06)
Analyst Disp		-0.205***	-0.203***		-0.208***	-0.208***
		(-4.45)	(-4.39)		(-4.50)	(-4.51)
IO ratio		-0.645***	-0.646***		-0.646***	-0.645***
		(-8.44)	(-8.45)		(-8.46)	(-8.44)
avg Ivol		3.269***	3.270***		3.268***	3.268***
		(40.63)	(40.64)		(40.61)	(40.61)
Firm-Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Day-Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	77697	76983	76983	77697	76983	76983
R-squared	0.522	0.574	0.574	0.522	0.574	0.574

Table 22: Efficient Analysts' Revisions and Implied Volatility

This table investigates the impact of efficient analysts' recommendation revisions on future uncertainty about a firm measured from option data. The table reports the regression result of option implied volatility on dummy variables of upgrade and downgrade revisions interacting with the *signal-to-noise volatility* ratio (SNR). log(ImpVol) is the log of option implied volatility. The dummy variables, *d_down* and *d_up*, are indicators equal to one if an observation is accompanied by a downgrade revision and an upgrade revision, respectively. $\log(SNR)$ is the log of the *signal-to-noise volatility* ratio, which measures stock price efficiency. $\log(ME)$, $\log(BM)$, *Amihud*, and Stock Spread(%) are the market capitalization, book-to-market ratio, illiquidity measure of [5], and percentage of bid-ask spread of underlying stock, respectively. Option related controls are the percentage of option bid-ask spread (Option Spread(%)), average of deltas of options (Delta, with put deltas being reversed in sign), analyst forecast dispersion (Analyst Disp), institutional ownership (IO ratio), and average of implied volatility (avg Ivol). All specifications below include the firm- and the day-fixed effects, and the standard errors are clustered at a daily level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Dependent Variable: log(ImpVol)					
	(1)	(2)	(3)	(4)	(5)	(6)
d_down	0.0236***	-0.0167***	0.201*			
	(4.98)	(-3.88)	(1.83)			
d_down× $\log(SNR)$			-0.0419**			
			(-1.98)			
d_up				0.0136***	-0.0225***	0.0567
				(2.99)	(-5.45)	(0.50)
d_up× $\log(SNR)$						-0.0152
						(-0.70)
$\log(ME)$		-0.0211***	-0.0211***		-0.0210***	-0.0210***
		(-7.92)	(-7.92)		(-7.91)	(-7.91)
$\log(BM)$		-0.00721***	-0.00724***		-0.00719***	-0.00720***
		(-4.21)	(-4.23)		(-4.20)	(-4.20)
$\log(vol)$		0.153***	0.153***		0.153***	0.153***
		(81.39)	(81.44)		(81.45)	(81.53)
Amihud		0.284***	0.284***		0.283***	0.283***
		(14.56)	(14.56)		(14.55)	(14.55)
Stock Spread(%)		0.00455	0.00455		0.00452	0.00452
		(1.57)	(1.57)		(1.56)	(1.56)
Option Spread(%)		-0.298***	-0.298***		-0.298***	-0.298***
		(-25.87)	(-25.87)		(-25.87)	(-25.87)
Delta		-0.417***	-0.417***		-0.416***	-0.416***
		(-25.28)	(-25.29)		(-25.25)	(-25.27)
Analyst Disp		0.249***	0.249***		0.250***	0.250***
		(12.30)	(12.28)		(12.27)	(12.27)
IO ratio		0.0908***	0.0909***		0.0910***	0.0909***
		(7.40)	(7.41)		(7.42)	(7.41)
Firm-Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Day-Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	82613	81871	81871	82613	81871	81871
R-squared	0.852	0.883	0.883	0.852	0.883	0.883

Table 23: Efficient Analysts' Revisions and the Slope of Implied Volatility Smile

This table investigates the impact of efficient analysts' recommendation revisions on the jump risk measured from option data. The table reports the regression result of the slope of implied volatility smile on dummy variables of upgrade and downgrade revisions interacting with the *signal-to-noise volatility* ratio (SNR). $Slope\_IVS$ is the slope of implied volatility smile constructed by running the regression of the out-of-the-money put option prices on their deltas of a firm each day. The dummy variables, *d_down* and *d_up*, are indicators equal to one if an observation is accompanied by a downgrade revision and an upgrade revision, respectively. $\log(SNR)$ is the log of the *signal-to-noise volatility* ratio, which measures stock price efficiency. $\log(ME)$, $\log(BM)$, *Amihud*, and Stock Spread(%) are the market capitalization, book-to-market ratio, illiquidity measure of [5], and percentage of bid-ask spread of underlying stock, respectively. Option related controls are the percentage of option bid-ask spread (Option Spread(%)), average of deltas of options (Delta, with put deltas being reversed in sign), analyst forecast dispersion (Analyst Disp), institutional ownership (IO ratio), and average of implied volatility (avg Ivol). All specifications below include the firm- and the day-fixed effects, and the standard errors are clustered at a daily level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Dependent Variable: $Slope\_IVS$					
	(1)	(2)	(3)	(4)	(5)	(6)
d_down	0.00000777	-0.0000483**	0.00219***			
	(0.37)	(-2.30)	(4.01)			
d_down×log(SNR)			-0.000430***			
			(-4.12)			
d_up				-0.00000200	-0.0000421**	0.000274
				(-0.09)	(-2.07)	(0.55)
d_up×log(SNR)						-0.0000609
						(-0.63)
log(ME)		-0.0000593***	-0.0000594***		-0.0000591***	-0.0000591***
		(-4.42)	(-4.43)		(-4.41)	(-4.41)
log(BM)		-0.000151***	-0.000152***		-0.000151***	-0.000151***
		(-16.38)	(-16.41)		(-16.36)	(-16.36)
log(vol)		0.000180***	0.000182***		0.000180***	0.000180***
		(20.94)	(21.07)		(20.94)	(20.95)
Amihud		0.00101***	0.00101***		0.00101***	0.00101***
		(16.60)	(16.60)		(16.57)	(16.58)
Stock Spread(%)		-0.00000660	-0.00000659		-0.00000663	-0.00000664
		(-0.41)	(-0.41)		(-0.41)	(-0.41)
Option Spread(%)		-0.000737***	-0.000738***		-0.000738***	-0.000739***
		(-11.69)	(-11.71)		(-11.69)	(-11.70)
Delta		-0.00135***	-0.00135***		-0.00135***	-0.00135***
		(-16.73)	(-16.77)		(-16.72)	(-16.73)
Analyst Disp		0.000799***	0.000795***		0.000800***	0.000801***
		(10.18)	(10.10)		(10.16)	(10.16)
IO ratio		0.000000652	0.00000196		0.00000165	0.00000140
		(0.01)	(0.04)		(0.03)	(0.03)
Firm-Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Day-Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	82613	81871	81871	82613	81871	81871
R-squared	0.559	0.581	0.581	0.559	0.581	0.581

Table 24: Subsample Analysis: Efficient Analysts' Recommendation Revision and Option Market Reaction

This table investigates the impact of informative recommendation revisions (downgrade and upgrade, separately) on option market reaction. The table reports the regression results of option market measures (O/S ratio, implied volatility, and slope of implied volatility smile) on dummy variables of upgrade and downgrade revisions interacting with the *signal-to-noise volatility* ratio (SNR) to examine the impact of informativeness of revisions on option prices. Panel A and Panel B show the results for the subsample with downgrade revisions and the subsample with upgrade revisions, respectively. To investigate the impact on the options market, three option market measures are used as dependent variables: $\log(O/S\,ratio)$ is the log of the ratio of option trading volume to stock trading volume, $\log(ImpVol)$ is the log of implied volatility smile, and $Slope\_IVS$ is the slope of implied volatility smile constructed by running the regression of the out-of-the-money put option prices on their deltas of a firm each day. $\log(SNR)$ is the log of the *signal-to-noise volatility* ratio, which measures stock price efficiency. $\log(ME)$, $\log(BM)$, $\log(vol)$, and Amihud are the market capitalization, book-to-market ratio, trading volume, and the illiquidity measure of [5], respectively. Option related controls are the percentage of option bid-ask spread (Option Spread(%)), average of deltas of options (Delta, with put deltas being reversed in sign), analyst forecast dispersion (Analyst Disp), institutional ownership (IO ratio), and average of implied volatility (avg Ivol). All specifications below include the firm-fixed effect, and the standard errors are clustered at a daily level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Panel A. Downgrade Revisions						Panel B. Upgrade Revisions					
	$\log(O/S\,ratio)$		$\log(ImpVol)$		$Slope\_IVS$		$\log(O/S\,ratio)$		$\log(ImpVol)$		$Slope\_IVS$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\log(SNR)$	0.412***	0.328***	0.0503	-0.119***	-0.000336*	-0.000663***	0.510***	0.290**	0.0815	-0.0607	0.0000630	-0.000362**
	(3.21)	(2.80)	(0.83)	(-2.74)	(-1.93)	(-4.21)	(3.78)	(2.39)	(1.54)	(-1.51)	(0.40)	(-2.57)
$\log(ME)$		0.922***		-0.504***		-0.00106***		0.942***		-0.366***		-0.000686***
		(9.24)		(-12.33)		(-8.44)		(9.68)		(-11.72)		(-5.87)
$\log(BM)$		-0.0361		-0.175***		-0.000473***		-0.0461		-0.109***		-0.000396***
		(-0.51)		(-8.06)		(-5.70)		(-0.66)		(-4.44)		(-4.45)
Amihud		0.347**		0.595***		0.00162***		0.0919		0.764***		0.00258***
		(2.42)		(4.02)		(3.95)		(0.45)		(6.25)		(9.02)
Stock Spread(%)		-0.569***		0.423***		0.000606***		-0.576***		0.424***		0.000827***
		(-5.30)		(9.53)		(3.45)		(-4.34)		(5.32)		(3.77)
Option Spread(%)		-1.860***		-0.207*		-0.00142***		-1.964***		-0.302***		-0.00104**
		(-5.75)		(-1.94)		(-3.35)		(-5.16)		(-2.71)		(-2.50)
Delta		-1.991***		-1.014***		-0.00558***		-1.621***		-1.385***		-0.00588***
		(-3.76)		(-5.42)		(-8.80)		(-2.72)		(-7.85)		(-9.46)
Analyst Disp		0.273		0.0137		-0.000502**		1.233***		0.274**		0.000497
		(1.58)		(0.20)		(-2.11)		(3.95)		(2.10)		(0.73)
IO ratio		-1.024		-0.366		-0.000104		-1.615**		-0.462*		-0.000500
		(-1.48)		(-1.63)		(-0.15)		(-2.34)		(-1.78)		(-0.60)
avgivol		0.709***						0.551*				
		(3.05)						(1.79)				
log(vol)				0.190***		0.000249***				0.124***		0.000241***
				(9.13)		(3.68)				(6.66)		(3.68)
Firm-Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1236	1231	1318	1313	1318	1313	1159	1154	1252	1246	1252	1246
R-squared	0.394	0.524	0.211	0.619	0.104	0.356	0.336	0.470	0.186	0.622	0.069	0.434

Table 25: Robustness Test: Excluding Earnings Announcements

Thjs table shows that our results of efficient analysts' recommendation revisions on stock and option markets are not driven by earnings announcements of a firm. We exclude firm-day observations if they are accompanied by earnings announcements. In Panel A, we report the regression result to show that the relationship between our measure of stock price efficiency, the SNR, and the release of downgrade or upgrade revisions is not driven by earnings announcements. In Panel B and Panel C, we show the robustness of the finding in stock and option markets, respectively. $\log(SNR)$ is the log of the *signal-to-noise volatility* ratio, which measures stock price efficiency. The dummy variables, *d_down* and *d_up*, are indicators equal to one if an observation is accompanied by a downgrade revision and an upgrade revision, respectively. $\log(ME)$, $\log(BM)$, $\log(vol)$, and Amihud are the market capitalization, book-to-market ratio, trading volume, and illiquidity measure of [5], respectively. In Panel C, option related controls are included: the percentage of option bid-ask spread (Option Spread(%)), average of deltas of options (Delta, with put deltas being reversed in sign), analyst forecast dispersion (Analyst Disp), institutional ownership (IO ratio), and average of implied volatility (avg Ivol). t-statistics are clustered at a daily level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Stock Price Efficiency and Recommendation Revisions

	Dependent Variable: $\log(SNR)$				
	(1)	(2)	(3)	(4)	(5)
d_down	0.0412***		0.0203***		0.0208***
	(7.73)		(3.95)		(4.04)
d_up		0.0363***		0.0176***	0.0182***
		(6.57)		(3.26)	(3.36)
$\log(ME)$			0.0104***	0.0103***	0.0105***
			(3.39)	(3.36)	(3.42)
$\log(BM)$			-0.0253***	-0.0254***	-0.0253***
			(-11.45)	(-11.47)	(-11.41)
$\log(vol)$			0.0836***	0.0837***	0.0832***
			(42.06)	(42.05)	(41.75)
Amihud			0.130***	0.131***	0.131***
			(13.72)	(13.72)	(13.73)
Firm-Fixed Effect	Yes	Yes	Yes	Yes	Yes
Day-Fixed Effect	Yes	Yes	Yes	Yes	Yes
Observations	81392	81392	81392	81392	81392
R-squared	0.222	0.222	0.252	0.252	0.252

Panel B. Efficient Analysts' Revisions and Stock Market Reaction

	Dependent Variable: AR(0, 0)									
	Downgrade Revisions					Upgrade Revisions				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
log(SNR)	-0.0124***	-0.00779***				0.00416**	0.00163			
	(-5.82)	(-3.71)				(2.01)	(0.77)			
log(SNR)×d_low			0.000468***					-0.0000355		
			(2.81)					(-0.21)		
log(SNR)×d_mid				0.0000897					-0.000122	
				(0.57)					(-0.73)	
log(SNR)×d_high					-0.000541***					0.000155
					(-3.26)					(0.93)
log(ME)		0.00244*	0.00254*	0.00253*	0.00261*		-0.00128	-0.00127	-0.00123	-0.00129
		(1.71)	(1.77)	(1.76)	(1.82)		(-0.74)	(-0.73)	(-0.71)	(-0.75)
log(BM)		0.00141	0.00165	0.00178	0.00154		-0.00219	-0.00226	-0.00222	-0.00218
		(1.08)	(1.27)	(1.36)	(1.18)		(-1.47)	(-1.52)	(-1.49)	(-1.46)
log(vol)		-0.00712***	-0.00747***	-0.00778***	-0.00724***		0.00558***	0.00566***	0.00568***	0.00558***
		(-8.19)	(-8.71)	(-9.19)	(-8.37)		(5.23)	(5.41)	(5.46)	(5.26)
Amihud		0.00765**	0.00805**	0.00808**	0.00743**		-0.000342	-0.000354	-0.000146	-0.000245
		(2.21)	(2.30)	(2.29)	(2.15)		(-0.10)	(-0.10)	(-0.04)	(-0.07)
Firm-Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1318	1318	1318	1318	1318	1252	1252	1252	1252	1252
R-squared	0.0544	0.122	0.119	0.115	0.121	0.0183	0.0563	0.0559	0.0561	0.0562

Panel C. Efficient Analysts' Revisions and Option Market Reaction

	Downgrade Revisions						Upgrade Revisions					
	$\log(O/S\,ratio)$		$\log(ImpVol)$		$Slope\_IVS$		$\log(O/S\,ratio)$		$\log(ImpVol)$		$Slope\_IVS$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\log(SNR)$	0.372***	0.289**	0.0716	-0.105**	-0.000244	-0.000558***	0.487***	0.237*	0.114**	-0.0379	0.000125	-0.000318**
	(2.75)	(2.35)	(1.20)	(-2.38)	(-1.41)	(-3.56)	(3.49)	(1.86)	(2.20)	(-0.92)	(0.83)	(-2.30)
$\log(ME)$		0.935***		-0.477***		-0.000985***		0.952***		-0.376***		-0.000712***
		(9.25)		(-12.58)		(-8.28)		(9.60)		(-12.12)		(-6.34)
$\log(BM)$		-0.0301		-0.197***		-0.000535***		-0.00826		-0.130***		-0.000451***
		(-0.41)		(-9.06)		(-6.38)		(-0.11)		(-5.19)		(-4.99)
Amihud		0.330*		0.681***		0.00188***		0.0864		0.685***		0.00240***
		(1.86)		(4.94)		(4.97)		(0.41)		(5.65)		(8.48)
Stock Spread(%)		-0.584***		0.420***		0.000590***		-0.547***		0.420***		0.000795***
		(-5.42)		(9.40)		(3.53)		(-4.29)		(5.30)		(3.70)
Option Spread(%)		-1.898***		-0.139		-0.00130***		-2.147***		-0.243**		-0.00113***
		(-5.64)		(-1.31)		(-3.07)		(-5.36)		(-2.13)		(-2.71)
Delta		-1.856***		-1.004***		-0.00552***		-1.543**		-1.515***		-0.00624***
		(-3.37)		(-5.40)		(-8.80)		(-2.45)		(-8.39)		(-10.47)
Analyst Disp		0.305*		-0.0601		-0.000587***		1.206***		0.216		0.000184
		(1.79)		(-0.82)		(-2.74)		(3.80)		(1.61)		(0.29)
IO Ratio		-1.105		-0.314		0.00000133		-1.797**		-0.556**		-0.000697
		(-1.52)		(-1.36)		(0.00)		(-2.54)		(-2.13)		(-0.86)
avg Ivol		0.832***						0.585*				
		(3.36)						(1.81)				
$\log(vol)$				0.207***		0.000251***				0.159***		0.000269***
				(9.73)		(3.68)				(7.84)		(4.10)
Firm-Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1162	1157	1240	1235	1240	1235	1100	1095	1183	1177	1183	1177
R-squared	0.389	0.521	0.216	0.638	0.107	0.375	0.328	0.464	0.184	0.619	0.0661	0.435

Table 26: Subsample Period Analysis: Stock Price Efficiency and Recommendation Revisions

This table presents the subsample period analysis results. In each panel of this table, we separate the sample period (2001-2014) into earlier (2001-2007) and later (2008-2014) subsample periods. In Panel A, we report the regression results on the relationship between the *signal-to-noise volatility* ratio (SNR) and the release of downgrade or upgrade revisions. In Panel B and Panel C, we provide the same results in stock and option markets for each subsample period, respectively. The dependent variable, log(SNR), is a measure of stock price efficiency. The dummy variables, *d_down* and *d_up*, are indicators equal to one if an observation is accompanied by a downgrade revision and an upgrade revision, respectively. log(ME), log(BM), log(vol), and Amihud are the market capitalization, book-to-market ratio, trading volume, and illiquidity measure of [5], respectively. In Panel C, option related controls are included: the percentage of option bid-ask spread (Option Spread(%)), average of deltas of options (Delta, with put deltas being reversed in sign), analyst forecast dispersion (Analyst Disp), institutional ownership (IO ratio), and average of implied volatility (avg Ivol). t-statistics are clustered at a daily level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Stock Price Efficiency and Recommendation Revisions

	Dependent Variable: log(SNR)											
	Subsample I. 2001 - 2007						Subsample II. 2008 - 2014					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
d_down	0.0558***	0.0565***	0.0186**			0.0190**	0.0376***	0.0385***	0.0199***			0.0205***
	(6.70)	(6.79)	(2.42)			(2.49)	(5.22)	(5.34)	(2.83)			(2.92)
d_up		0.0436***		0.0427***	0.0189**	0.0193**		0.0497***		0.0490***	0.0282***	0.0287***
		(5.19)		(5.07)	(2.42)	(2.48)		(6.79)		(6.70)	(3.87)	(3.94)
log(ME)			-0.00670		-0.00664	-0.00613			-0.00397		-0.00371	-0.00426
			(-0.95)		(-0.94)	(-0.87)			(-0.48)		(-0.45)	(-0.51)
log(BM)			-0.131***		-0.131***	-0.131***			-0.0429***		-0.0424***	-0.0428***
			(-28.61)		(-28.65)	(-28.57)			(-7.07)		(-6.99)	(-7.06)
log(vol)			0.0712***		0.0712***	0.0707***			0.0678***		0.0676***	0.0671***
			(19.15)		(19.09)	(18.93)			(21.40)		(21.28)	(21.14)
Amihud			0.961***		0.962***	0.962***			-0.0132**		-0.0132**	-0.0131**
			(18.08)		(18.06)	(18.08)			(-2.11)		(-2.12)	(-2.11)
Firm Fixed-Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed-Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	40456	40456	40456	40456	40456	40456	42171	42171	42171	42171	42171	42171
R-squared	0.054	0.055	0.163	0.053	0.163	0.163	0.057	0.058	0.087	0.057	0.088	0.088

Panel B. Efficient Analysts' Revisions and Stock Market Reaction

	Dependent Variable: AR(0, 0)											
	Subsample I. 2001 - 2007						Subsample II. 2008 - 2014					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
d_down	-0.00658***	0.0496***			0.0515***		-0.00722***	0.0642***			0.0677***	
	(-9.80)	(3.07)			(3.17)		(-12.96)	(4.38)			(4.57)	
d_down×log(SNR)		-0.0108***			-0.0112***			-0.0137***			-0.0145***	
		(-3.45)			(-3.55)			(-4.85)			(-5.06)	
d_up			0.00750***	-0.00130		-0.00119			0.00666***	-0.0365**		-0.0349**
			(11.55)	(-0.08)		(-0.08)			(10.95)	(-2.27)		(-2.19)
d_up×log(SNR)				0.00170		0.00169				0.00827***		0.00794***
				(0.55)		(0.55)				(2.68)		(2.58)
log(ME)					0.000356	0.000743**					0.00147***	0.00119***
					(1.02)	(2.14)					(3.67)	(2.95)
log(BM)					-0.000383	-0.000194					0.000161	0.0000551
					(-1.47)	(-0.74)					(0.40)	(0.14)
log(vol)					0.000255	-0.000148					0.00113***	0.000748***
					(1.25)	(-0.73)					(5.67)	(3.74)
Amihud					0.00115	0.00125					0.000208	0.000262
					(0.50)	(0.54)					(0.34)	(0.43)
Firm Fixed-Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	40456	40456	40456	40456	40456	40456	42120	42120	42120	42120	42120	42120
R-squared	0.005	0.005	0.006	0.006	0.006	0.006	0.007	0.008	0.006	0.006	0.010	0.007

Panel C. Efficient Analysts' Revisions and Option Market Reaction

	Subsample I. 2001 - 2007						Subsample II. 2008 - 2014					
	log($O/S\,ratio$)		log($ImpVol$)		$Slope\_IVS$		log($O/S\,ratio$)		log($ImpVol$)		$Slope\_IVS$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
d_down	-1.137 (-1.28)		0.154 (0.66)		0.00128 (1.15)		-1.639** (-2.14)		1.577*** (6.17)		0.00422*** (5.32)	
d_down×log (SNR)	0.236 (1.40)		-0.0268 (-0.59)		-0.000225 (-1.05)		0.332** (2.27)		-0.312*** (-6.37)		-0.000837*** (-5.52)	
d_up		-0.298 (-0.34)		-0.306 (-1.50)		-0.000478 (-0.55)		-2.122*** (-2.65)		1.309*** (4.01)		0.00259*** (2.59)
d_up×log (SNR)		0.0796 (0.48)		0.0544 (1.38)		0.0000838 (0.50)		0.424*** (2.77)		-0.257*** (-4.11)		-0.000501*** (-2.65)
log(ME)	0.552*** (20.15)	0.202*** (7.35)	-0.312*** (-30.37)	-0.313*** (-30.46)	-0.000778*** (-23.23)	-0.000782*** (-23.24)	1.099*** (28.42)	1.100*** (28.42)	-0.674*** (-44.44)	-0.675*** (-44.51)	-0.000807*** (-16.27)	-0.000811*** (-16.34)
log(BM)	-0.131*** (-5.89)	-0.426*** (-19.31)	-0.186*** (-35.22)	-0.186*** (-35.38)	-0.000397*** (-21.16)	-0.000398*** (-21.28)	0.136*** (5.01)	0.139*** (5.09)	-0.175*** (-16.08)	-0.176*** (-16.19)	0.000195*** (5.03)	0.000192*** (4.94)
Amihud	-0.260* (-1.72)	-0.885*** (-6.69)	2.362*** (28.06)	2.360*** (28.05)	0.00458*** (17.98)	0.00458*** (17.93)	0.109** (2.56)	0.109** (2.57)	0.406*** (14.92)	0.406*** (14.93)	0.00153*** (17.70)	0.00153*** (17.71)
Stock Spread(%)	-0.310*** (-15.27)	-0.147*** (-8.21)	0.188*** (31.19)	0.188*** (31.16)	0.000135*** (9.68)	0.000134*** (9.64)	-0.263*** (-3.03)	-0.262*** (-3.03)	0.376*** (4.67)	0.376*** (4.67)	0.000977*** (4.42)	0.000978*** (4.42)
Option Spread(%)	-0.378*** (-3.62)	-0.591*** (-6.18)	-0.881*** (-31.38)	-0.880*** (-31.36)	-0.00170*** (-17.02)	-0.00170*** (-16.97)	-0.757*** (-7.00)	-0.756*** (-7.00)	0.0254 (0.61)	0.0249 (0.60)	-0.00202*** (-15.15)	-0.00203*** (-15.17)
Delta	0.206 (1.38)	-0.713*** (-5.21)	-1.624*** (-41.66)	-1.623*** (-41.62)	-0.00374*** (-25.92)	-0.00374*** (-25.88)	0.0813 (0.49)	0.0771 (0.47)	-0.162*** (-3.09)	-0.159*** (-3.04)	-0.00456*** (-25.53)	-0.00456*** (-25.48)
Analyst Disp	-0.625*** (-2.85)	0.448** (2.39)	0.405*** (6.43)	0.404*** (6.42)	-0.000287 (-1.61)	-0.000299* (-1.67)	0.414*** (7.43)	0.411*** (7.40)	-0.160*** (-5.93)	-0.158*** (-5.85)	-0.000339*** (-3.40)	-0.000333*** (-3.34)
IO ratio	0.463*** (2.93)	-0.423*** (-2.85)	-1.019*** (-35.17)	-1.019*** (-35.16)	-0.000978*** (-7.66)	-0.000980*** (-7.67)	-0.335*** (-2.79)	-0.332*** (-2.76)	0.248*** (7.56)	0.246*** (7.49)	0.000817*** (8.63)	0.000813*** (8.59)
avgivol	1.934*** (17.20)	4.887*** (34.76)					0.629*** (7.77)	0.627*** (7.76)				
log(vol)			0.0910*** (17.31)	0.0919*** (17.47)	0.0000588*** (3.72)	0.0000623*** (3.93)			0.247*** (29.53)	0.245*** (29.43)	0.000433*** (16.90)	0.000428*** (16.79)
Firm Fixed-Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	36954	36954	40099	40099	40099	40099	40031	40031	41772	41772	41772	41772
R-squared	0.320	0.441	0.662	0.662	0.266	0.266	0.557	0.557	0.732	0.732	0.460	0.460

Table 27: Influential Analyst vs. Efficient Analyst

This table investigates whether the impact of efficient analysts is permanent or transitory by examining dynamics of abnormal returns on and after the date of revision issuance. For each firm in the sample, we construct quartile portfolios based on log(SNR). Panel A and Panel B report average cumulative abnormal returns for each quartile for upgrade and downgrade revisions, respectively.

Panel A. Upgrade revisions				
(1) Abnormal return on day of revision				
		On SNR		
		1 (Low)	2	3 (High)
On Daily	1 (Low)	-1.635%	-1.366%	-1.568%
Return	2	0.019%	0.004%	-0.024%
	3 (High)	1.703%	1.627%	2.047%
(2) CAR(1,5)				
		1 (Low)	2	3 (High)
On Daily	1 (Low)	0.804%	0.108%	-0.417%
Return	2	-0.057%	0.175%	0.649%
	3 (High)	0.054%	-0.001%	0.416%
(3) CAR(1,10)				
		1 (Low)	2	3 (High)
On Daily	1 (Low)	1.336%	-0.403%	-0.289%
Return	2	-0.142%	0.416%	0.809%
	3 (High)	0.090%	0.221%	0.548%
Panel B. Downgrade revisions				
(1) Abnormal return on day of revision				
		On SNR		
		1 (Low)	2	3 (High)
On Daily	1 (Low)	-1.950%	-1.849%	-2.264%
Return	2	-0.093%	-0.113%	-0.088%
	3 (High)	1.214%	1.357%	1.432%
(2) CAR(1,5)				
		1 (Low)	2	3 (High)
On Daily	1 (Low)	-0.330%	-0.438%	0.178%
Return	2	-0.165%	-0.414%	-0.081%
	3 (High)	0.249%	-0.745%	-0.686%
(3) CAR(1,10)				
		1 (Low)	2	3 (High)
On Daily	1 (Low)	-0.301%	-0.359%	0.352%
Return	2	-0.411%	-0.116%	0.163%
	3 (High)	0.347%	-0.476%	-0.390%

CHAPTER III

REALIZED SKEWNESS FOR INFORMATION AMBIGUITY

3.1 Introduction

Investors process financial market news, or signals, every day. When there is incomplete knowledge about signal quality, and hence when the quality of information is difficult to judge, investors face ambiguous information. When interpreting such ambiguous news, investors do not know the exact probability distribution to describe the relevant information. The impact of ambiguous information on various financial decisions has been proven important in many recent studies (for example, [51], [91], and [77] among others). In this paper, we empirically investigate the effect of ambiguous information on investors' responses and its implications on asset prices. Specifically, we focus on the asymmetric response of investors to the releases of ambiguous information and its association with subsequent stock returns.

If the exact distribution of the signals is not known due to the ambiguity of information, investors dislike the situation (i.e., become averse) and are expected to respond asymmetrically to the releases of such information. This is because ambiguity-averse investors behave as if they take the worst-case assessment of the ambiguous information. For example, if an investor views the ambiguous information as good news, the worst case scenario is that the information is unreliable. On the other hand, if an investor views the ambiguous information as bad news, the worst case scenario is that the information is reliable. Given this asymmetric assessment, the investor's response to ambiguous information releases becomes asymmetric.[1] The greater the

[1] [51] provide a theoretical framework showing how ambiguous information generates asymmetric

degree of information ambiguity, the distribution of investor response would be more negatively skewed.

Given the aforementioned intuition behind the negative skewness, we propose to use realized skewness constructed using high frequency data to assess the ambiguity of information. The use of intraday data is important because it allows us to precisely capture the asymmetric response of investors to the release of ambiguous information, which can be much weaker or even absent when using daily data. We confirm the negatively skewed stock return distribution around the ambiguous information releases. In particular, we examine how realized skewness varies around the releases of analyst earnings forecasts and recommendation revisions, which are two main examples of ambiguous (or intangible) information. We document that average realized skewness on days with ambiguous information releases is significantly lower than the average realized skewness on days without those information releases. To the best of our knowledge, this is the first paper that empirically documents that realized skewness of individual stock returns is associated with the information ambiguity.

Our empirical analyses reveal that our realized skewness measure, as a proxy for information ambiguity, has a strong predictive power in explaining subsequent stock returns around the information releases. The literature on return continuation due to investors' under-reactions documents a strong drift of stock prices around news events. Similarly, we find that realized skewness predicts subsequent five, ten, and twenty-day cumulative returns after earnings forecasts and analyst recommendation releases. Our finding implies that the initial reaction of investors to news releases is incomplete, especially when the information is ambiguous, resulting in the short-term return predictability. We interpret that investors require more time and effort to quantify the impact of information for their financial decision making when information is more ambiguous. This evidence is consistent with [109] which documents

---

stock return distribution.

stronger return continuation around information releases when information ambiguity about a firm is more prominent. However, unlike [109], we first focus on asymmetric responses of investors generated by ambiguous information releases, and further relate the asymmetric responses to subsequent return dynamics. Furthermore, we use high-frequency data to capture more frequent time variation of information ambiguity in relevant markets because information is much quickly incorporated into stock prices in recent years due to advances in information technology ([27]).

The finding on realized skewness explaining the pattern of subsequent stock returns around information releases is unique in that the existing evidence on return reversal or continuation is not able to explain our finding. The vast literature on short-term return predictability documents evidence that a large price movement tends to exhibit a drift when the movement is accompanied by the news releases, while it tends to exhibit a reversal when there are no relevant information releases. We find that a return continuation (i.e., a positive predictability of contemporaneous stock returns in explaining subsequent returns) does not exist once we account for information ambiguity proxied by realized skewness. That is, while realized skewness strongly predicts a pattern of subsequent stock returns, a contemporaneous return does not play any role in predicting subsequent returns.

As more negatively skewed stock returns imply greater information ambiguity, we further divide our sample into positive and negative skewness subsamples. From the subsample analysis, we find that the predictive power of realized daily skewness is significant for the negative realized skewness sample, while the predictive power is insignificant for the positive skewness sample. This finding suggests that stocks with greater negative realized skewness experience stronger under-reaction by investors, yielding lower returns in subsequent periods. However, stocks with less information ambiguity (with positive realized skewness) experience a much lower degree of under-reaction, resulting in insignificant return predictability.

We verify the economic significance of the main result of this paper by demonstrating the superior performance of a trading strategy that exploits our finding on skewness. The main results of this paper suggest two different aspects of stock return predictability, both of which can be used in investment strategies. First, realized skewness is able to predict subsequent returns on days with the release of ambiguous information due to investors' under-reactions. Second, a contemporaneous stock return is able to predict subsequent returns on days without such information due to short-term return reversals. These two distinctive patterns can be exploited independently because the contemporaneous variables in each case are different: realized skewness or return. Thus, we provide a zero-net investment trading strategy that exploits both features.

In our zero-net investment trading strategy, one portfolio is constructed with stocks experiencing the release of analyst earnings forecasts or recommendations over the last week of each month, and the other portfolio includes all other stocks without this information over the same week. For the first portfolio with information, stocks are further categorized into a positive realized skewness portfolio and a negative realized skewness portfolio to exploit investors' under-reactions due to information ambiguity.[2] For the second portfolio, stocks are also separated into a negative return portfolio and a positive return portfolio to exploit the return reversals on no-information days.[3] The zero-net investment portfolio combines these four portfolios by taking long positions on a positive realized skewness portfolio and a negative return portfolio while taking short positions on a negative realized skewness portfolio and a positive return portfolio.

The performance of our zero-net investment trading strategy with a Sharpe ratio

[2] We use an average of realized skewness over the last week of each month to categorize stocks into a positive or a negative realized skewness portfolio.

[3] We use an average of daily returns over the last week of each month to categorize stocks into a positive or a negative return portfolio.

of 1.766 is superior to other well-known zero-net investment portfolios, such as market portfolio (in excess of risk-free rate), size portfolio, value portfolio, and momentum portfolio. In particular, the inclusion of our finding on realized skewness increases the Sharpe ratio by 57% relative to that of the return reversal strategy (1.126). The return of the zero-net investment strategy is not exposed to common risk factors (the market, size, value, and momentum factors). The alpha of the strategy after taking the [30] four factors into account is 86.77 bps per month.

The other possible approach to investigate information ambiguity at a frequency comparable to our analysis is to use well-known proxies in the literature on information asymmetry,[4] because the release of information considered in this paper is strongly associated with the changes in the proxies for asymmetric information. Using various proxies for information asymmetry, such as the bid-ask spread, idiosyncratic volatility, illiquidity measure, and change in turnover ratio – all of which are widely believed to capture some degree of information ambiguity at a daily or higher frequency – we show that realized skewness shares a similar aspect of information ambiguity with other existing proxies. In addition to the theoretical guidance that realized daily skewness can be considered a proxy for information ambiguity, we find that realized skewness has the unique ability to explain return continuations not possessed by other existing proxies for information ambiguity. In particular, an orthogonalized component of realized skewness that eliminates common components explained by existing proxies (orthogonalized realized skewness) remains statistically significant in explaining subsequent returns under severe information ambiguity.

The present paper contributes to several strands of literature. First, our finding on realized skewness extends the recent studies on realized higher moments by showing that higher moments can be a useful tool to better understand the rapidly changing

[4]For example, see [106] and [63] for the bid-ask spread, [26] for liquidity, [46] for trading volume, and [84] for idiosyncratic volatility. Further discussion is provided in Section 3.5.

environment of the financial market. [4] is one of a few papers that studies realized higher moments using high-frequency data. They find that the *weekly* average of realized daily skewness is negatively priced in the cross-section of stock returns, which is explained by the skewness preference of investors. Different from their study documenting weekly realized skewness as a risk factor, our study views realized skewness as a measure of information ambiguity at a higher frequency.[5] [72] document that negative skewness tends to be more pronounced for stocks with larger market capitalization than those with smaller market capitalization. One of the potential reasons is that, as noted in [51], large stocks are covered in the news media more frequently than small stocks, resulting in a significant amount of ambiguous information to investors. This paper aims to investigate a different aspect of information ambiguity. By focusing on the events generating the influx of ambiguous information, we are able to capture the time variation of information ambiguity about firm fundamentals.

Second, this paper empirically contributes to the literature on information ambiguity by proposing a proxy to assess information ambiguity. [51] theoretically show the relation of ambiguous information and expected stock returns, and [77] examines the interaction between risk and information ambiguity and its effect on optimal portfolios. We complement the literature not only by developing a measure to empirically proxy information ambiguity but also by examining the impact of information ambiguity using realized skewness on investor behavior and subsequent stock return dynamics.

Third, this paper provides evidence that realized skewness can be a proxy for information ambiguity concerning a firm. Although information ambiguity generates significant frictions in financial markets, few proxies measure information ambiguity with clear directional interpretation and theoretical guidance. Realized skewness has

[5] In particular, we use an average of daily realized skewness over a three-day window centered on the day of ambiguous information release.

advantages over exiting proxies due to rich information on intra-day return dynamics and distribution that are not available when using other existing proxies, such as the bid-ask spread, illiquidity, change in the turnover ratio, or idiosyncratic volatility.[6]

Fourth, we contribute to the literature on return predictability by showing that realized skewness as a measure of information ambiguity predicts subsequent stock returns. While [109] finds that greater information ambiguity generates stronger price drifts, [51] predict that ambiguity-averse investors require compensation for information ambiguity in the form of higher expected returns. Our finding supports that of [109] in a sense that greater information ambiguity, as measured by realized skewness computed with high-frequency data, predicts lower subsequent returns by generating stronger under-reaction of investor. To the best of our knowledge, the present paper is the only paper to both demonstrate that realized skewness indeed captures the information ambiguity of a firm, and that realized skewness can play a role in refining our understanding of return predictability around news releases.

The remainder of the paper is organized as follows. Section 3.2 describes the sample data and provides summary statistics, Section 3.3 presents the main finding of the paper, Section 3.4 shows the economic significance of main findings by providing a profitable trading strategy, and Section 3.5 examines the relationship of realized skewness with other existing proxies for information ambiguity. In Section 3.6, we demonstrate the robustness of the main finding, and Section 3.7 concludes.

3.2 Data and Sample Description

Our main objective is to capture asymmetric responses of investors due to ambiguous information releases. As stock prices incorporate newly released information much quickly by virtue of technological advances in financial markets ([27]), high-frequency

[6] For example, see [40] and [64] for the bid-ask spread, [5] and [100] for liquidity, and [37] and [93] for idiosyncratic volatility regarding their informational content. Further discussion is provided in Section 3.5.

data is ideal to capture the responses at a high frequency such as a daily basis. Thus, in our empirical analysis, we investigate all stocks listed in the Trade and Quote (TAQ) database. The sample period for our analysis is from January 2, 2001 to May 30, 2014. We consider this recent sample period to mitigate concerns regarding infrequent trading and illiquidity problems that may contaminate our main findings. To calculate higher moments, we record the prices of all stocks in the TAQ database at five-minute intervals from 09:30 EST until 16:00 EST, and construct the five-minute returns as the difference between log prices with five-minute marks, as in [8]. To filter out stocks with infrequent trading, we require a stock to have at least 100 transactions per day[7] and use the recorded prices close to the five-minute interval time grid. In addition to these requirements, we further exclude stocks that have a closing price of less than five dollars.

In the present paper, we consider analyst earnings forecasts and recommendation releases as main examples of ambiguous news about a firm's future cash flows. Because the chosen information events are firm specific, we collect all firm-day observations for both events from the International Brokers Estimation System (I/B/E/S) database from January 2001 to May 2014.[8] Several studies on analyst recommendations and earnings forecasts document that these information events tend to occur overnight.[9] When firm-related events occur during the day, analysts usually examine the issues during business hours and release their reports on the same day or one day after the event. In certain cases, analysts can release their reports on a firm before the issue is

[7] Different thresholds for the minimum number of transactions within a day do not alter the main findings of our paper.

[8] Although we use high-frequency data to calculate realized higher moments, our analysis does not depend on time stamps at the intra-day level for these information releases. This is because our study considers how the realized moments of stock returns accompanied by information releases at the daily level affect returns over the next 5 to 20 trading days. Hence, the intra-day time-stamp delay concern that [23] raise is not an issue for our study.

[9] [3] report that the percentage of overnight recommendation releases is 61%. [23] report that this percentage is over 70% and earnings releases occur overnight in over 80% of cases after the year 2003.

released to the public in order to offer their prior expectations. For all these reasons, when we identify the dates of information releases, we cover one day before and after the release dates recorded in the I/B/E/S database.

These firm-specific news events are widely considered as the most influential information events, which contain a significant amount of intangible information about a firm's future cash flow.[10] This choice is also motivated by the finding that return dynamics are significantly altered in response to information arrivals (see [7]). The other reason that we select these two types of information is that they cover most firm-related news for a given firm. While earnings forecasts provide valuable information on cash flow, analyst reports offer an extensive coverage of a wide array of information (see [102]). Thus, by including not only earnings forecasts but also analyst reports for a firm, we expect to consider a comprehensive set of information about the firm.

We obtain data on market capitalizations, trading volumes, and daily bid-ask spreads from the Center for Research in Security Prices (CRSP) database. Accounting data, such as the book values of individual firms, are obtained from the COMPUSTAT database. For returns over horizons beyond one trading day, we use daily returns from the CRSP database for corresponding firms and dates instead of using high-frequency returns from the TAQ database. Given all these filtering requirements, the total number of companies covered in our sample varies, ranging from 1,748 to 3,903 per year, depending on changes in market conditions over the sample period.

We measure the asymmetric response at a daily level using *realized daily skewness*[11] computed with high-frequency returns. To control for other effects of realized

[10] As a robustness check, we also consider analyst recommendations and earnings forecasts separately to distinguish between the effects of the two types of information and find that analyst recommendation reports are the main driver of our results. However, including both types of information strengthens the main findings of our paper, which indicates that it is impossible to perfectly distinguish between the effects of these information events because of confounding event times between the two.

[11] [95] proposes an alternative measure of realized skewness. Because our main objective is not

higher moments, we include realized daily volatility and kurtosis in our main analysis. [4] use these measures to examine the relationship between realized higher moments and the subsequent week's returns.

We define the five-minute log return as the difference between log prices observed at five-minute intervals. A l-th intra-day return for a k-th firm on day t is first constructed by

$$r_{k,t,l} = \log P_{k,t,l+1} - \log P_{k,t,l}, \tag{24}$$

where $P_{k,t,l}$ is a l-th intra-day price of the k-th firm observed on day t. From the five-minute log returns obtained above, our measure of information ambiguity, realized daily skewness ($RDSkew_{k,t}$) for the k-th firm on day t, is computed as the sum of cubed high-frequency returns standardized with realized daily variance ($RDVar_{k,t}$):

$$RDSkew_{k,t} = \frac{\sqrt{n}\sum_{l=2}^{n} r_{k,t,l}^{3}}{RDVar_{k,t}^{3/2}}[12], \tag{25}$$

where n is the number of intra-day return observations in a day, and

$$RDVar_{k,t} = \sum_{l=1}^{n} r_{k,t,l}^{2}.[13] \tag{26}$$

Since we record five-minute prices from 09:30 ETS to 16:00 ETS, we have $n = 78$ for each day. As one of our control variables, we also use realized daily kurtosis ($RDKurt_{k,t}$) for the k-th firm on day t, which is defined as

$$RDKurt_{k,t} = \frac{n\sum_{l=2}^{n} r_{k,t,l}^{4}}{RDVar_{k,t}^{2}}. \tag{27}$$

to obtain a true measure of realized skewness but rather to examine the relationship between the asymmetric distribution of high-frequency returns and future stock returns, we do not follow his approach. In addition, the approach that [95] proposes is not applicable to our analysis because our primary objective is to capture the effect of daily-level information on asymmetries in high-frequency returns. As the methodology of [95] employs daily options data to estimate a lower-frequency measure of realized skewness (for example, weekly or monthly realized skewness), it is not appropriate for the purpose of our analysis.

[12] [4] use weekly measures of higher moments by taking averages of daily moments and demonstrate the predictability of realized skewness in the next week's returns and the robustness of the realized higher moment inference to microstructure noise. As our main objective is different from theirs and focuses on realized skewness as a proxy for information ambiguity concerning a firm's fundamentals, we use a daily measure of realized skewness using five-minute returns to precisely capture the impact of information.

[13] Realized daily volatility, $RDVol_{k,t}$, can be simply computed as $RDVol_{k,t} = \sqrt{RDVar_{k,t}}$.

It is inevitable that high-frequency returns contain microstructure noise and that realized moments on a daily basis computed from high-frequency returns are also contaminated with such noise. In fact, it is not the main objective of this paper to separate these components. However, a five-minute grid is a reasonable choice for the optimal sampling frequency that optimizes the trade-off between bias and efficiency gain in the estimation of realized moments (see, for example, [13, 14]). Using a Monte Carlo simulation, [4] show that realized daily moments from a finite sample are well behaved. Thus, despite the microstructure noise embedded in high-frequency data, it seems safe to implicitly assume that realized daily skewness captures the asymmetry of the distribution of intra-day returns.

Table 1 presents the time-series summary statistics of the annual means and medians for the main variables used in this paper. This table shows that the number of firms in our sample decreased significantly during the period of the recent financial crisis. For example, in 2005, the number of stocks in our sample is 3,903, while this number decreases to 1,748 in 2009. We also observe a considerably higher average (median) realized daily volatility of 3.4% (2.8%) in 2008 when compared to 1.8% (1.6%) in the years 2006 or 2007. As realized daily volatility shows a clear time-series pattern that depends on market conditions, realized daily skewness also captures time-varying properties of returns. While average realized daily skewness typically remains positive during normal market conditions, it became negative during the recent financial crisis. Realized daily kurtosis clearly exhibits higher average values in the early period of the sample (2001-2003) and around the recent financial crisis (2007) than in other periods. Overall, realized daily moments (variance, skewness, and kurtosis) display similar time variations to the sample moments. The other control variables for our regression analysis, such as the book-to-market ratio and the ratio (in percentage) of trading volume to total shares outstanding, also exhibit time variations, as expected.

To examine how ambiguous news releases (analyst earnings forecasts and recommendations) affect realized skewness of a firm, which leads to differential cross-sectional variations in subsequent stock returns, we create subsamples with and without these firm-specific information releases. Table 29 presents the time-series summary statistics (by year) for all firm-day observations in our full sample, observations with earnings forecasts, observations with analyst recommendation releases, observations with either one of two information releases, and observations without any of these information releases. For each year, we list the total number of all firm-trading day observations and average firm sizes in each of the five categories. We have more than 6 million firm-day observations in our full sample. We then separately report those statistics for the subsamples accompanied by earnings forecasts and analyst recommendation releases. Because we are not interested in the separate impacts of earnings forecasts and analyst recommendations on the relationship between realized skewness and subsequent stock returns, we combine the two subsamples and call the resulting sample the *Information sample* and the rest of the sample the *No-information sample*. For each sample, we now take an average of realized daily skewness over three days in order to minimize noise in high-frequency data and to be consistent with the identification of information events in I/B/E/S data. The fact that the impacts of earnings forecasts and analyst recommendation releases can be confounded with one another is not a concern in this study, as we are interested in how these types of intangible (ambiguous) information affect the (higher moments of) stock return distribution and the degree of return predictability in a subsequent period. From the descriptive statistics in Table 29, it is noteworthy that the *Information sample* includes larger firms than the *No-information sample*[14] with the average size in the *Information sample* being \$11.7 billion, while that for the *No-information*

[14]The reason that the *Information sample* has larger-sized firms than the *No-information sample* is largely that analysts usually follow firms with large market capitalization.

sample is \$5.6 billion.

3.3 Empirical Findings

3.3.1 The Impact of Ambiguous Information on Realized Moments

In this subsection, we study the impact of ambiguous information releases on realized daily moments in order to observe how our proposed measure of information ambiguity, realized skewness, behaves in response to information arrivals.

Table 30 reports the averages of realized daily higher moments for the full sample (*Full*), the subsamples accompanied by either earnings forecasts (*Forecast*) or analyst recommendations (*Recommendation*), the *Information sample* (*Information*), and the *No-information sample* (*No-information*). The differences in the measures between the *Information sample* and the *No-information sample* are reported in the last column. Realized daily skewness decreases significantly when earnings forecasts or recommendation reports are released, while realized volatility and realized kurtosis increase. The strong statistical significance of these differences suggests that these information releases affect the distribution of high-frequency returns, implying that realized daily moments are directly related to a firm's informational environment.

[51] theoretically show that ambiguity of information generates skewness in stock returns and that returns become more negatively skewed as information is more ambiguous. Their concept of ambiguity depends on the importance of intangible information (for example, analyst recommendations, earnings forecasts, or media reports) relative to tangible information (for example, dividend announcements). In line with the concept of ambiguity in [51], days with information (earnings forecasts and analyst recommendations) releases are accompanied by more intangible information than days without these information releases. In Table 30, we find evidence that realized daily skewness in the *Information sample* decreases significantly compared to the *No-information sample*, consistent with the argument in [51] that greater information

ambiguity leads to more negative skewness of stock returns.

3.3.2 Double-sorting on Contemporaneous Return and Realized Skewness

To examine the specific role of realized skewness as a measure of information ambiguity, we investigate its relationship with subsequent stock returns, depending on the information releases. For this analysis, it is crucial to distinguish the relationship from well-known factors of return reversals and continuations. For example, the literature on return reversals and continuations shows that price changes on days with information releases (for example, earnings announcements, headline news, or analyst reports) are followed by drift, while those on no-information days tend to reverse (see, for example, [33] and [102], among many others). Thus, in this subsection, we document the behavior of realized skewness depending on information releases and the impact of realized skewness on subsequent returns while accounting for return reversals and continuations.

To examine the role of realized skewness in explaining subsequent returns while controlling for contemporaneous returns, we first double-sort stocks in each sample based on contemporaneous returns and realized daily skewness. That is, for each day, we sort stocks based on their contemporaneous returns ($Ret_{k,t}$) into tercile portfolios. Within each of these terciles, stocks are sorted into tercile portfolios based on realized daily skewness ($RDSkew_{k,t}$). All portfolios are constructed with equal weights.[15] All portfolio returns and return differences of the top terciles and the bottom terciles (*H-L*) are exhibited in basis points. These double-sorted portfolios based on contemporaneous returns and realized daily skewness are constructed daily using the *Full sample*, the *Information sample*, and the *No-information sample*, and their next 5-, 10-, and 20-day cumulative returns are reported in Panel A, Panel B, and Panel C of Table 31, respectively.

[15]All results are robust to a value-weigthing scheme.

In Panel A of Table 31, the negative and significant return differences between the highest contemporaneous return portfolios and the lowest contemporaneous return portfolios (*H*-*L*) for all subsequent 5-, 10-, and 20-day returns show the prevalence of return reversals. The impact of realized skewness is statistically significant only in the highest contemporaneous return terciles. When we focus on the *Information sample* in Panel B, we find a statistically significant and positive relationship between realized skewness and subsequent stock returns. The statistical significance is shown in the highest and the lowest terciles of contemporaneous returns: the return difference between the highest and the lowest realized skewness terciles among the lowest (the highest) contemporaneous return stocks is 10.60 bps (10.69 bps) over the next 20 trading days. As expected, in Panel C, we find the evidence of return reversals and no strong evidence of realized skewness in explaining subsequent returns when we focus on days without an influx of ambiguous information.

Overall, the findings in Table 31 imply that realized daily skewness is an important factor in explaining the under-reaction of investors on information days as realized daily skewness absorbs return continuations in the *Information sample*. While the double-sorting analysis reveals a univariate relationship between realized skewness and future returns after controlling for contemporaneous return, we examine a multivariate relationship using the [58] regression in the next subsection.

3.3.3 Fama-MacBeth Regression Result

In order to test the impact of information ambiguity on subsequent stock returns while controlling for other existing factors that affect stock returns, we estimate the following specification using the [58] regression:[16]

$$Ret_{k,t+i,t+j} = \alpha + \beta_1 RDSkew_{k,t} + \beta_2 RDVol_{k,t} + \beta_3 RDKurt_{k,t} + \gamma' X_{k,t} + u_{k,t+i,t+j}, \quad (28)$$

where $Ret_{k,t+i,t+j}$ is the cumulative return (in bps) of the k-th stock over a period from day $t+i$ to day $t+j$, and realized higher moments, $RDSkew_{k,t}$, $RDVol_{k,t}$, and $RDKurt_{k,t}$, of the k-th stock returns on day t are computed using equations (25), (26), and (27). $X_{k,t}$ represents a vector of several control variables for the k-th firm observed at the end of day t, and $u_{k,t+i,t+j}$ is an error term.

As discussed in the previous section, one of the major concerns associated with using regression specification (28) are the well-known return reversals and continuations. Recently, [102] documents that large price changes accompanied by information releases are followed by return continuations, while those with no information result in reversals. [97], [33], and [105] also obtain a similar empirical finding using different sets of information. Thus, to control for the existing evidence on short-term return predictability, we include the return (in bps) of firm k on day t, $Ret_{k,t}$, in all regressions as one of the main control variables.

Because it is well documented in the literature that size, book-to-market, and momentum predict the cross-section of stock returns ([54], [56], [83]), we include a log of market size ($LogME_{k,t}$), a log of book-to-market ratio ($LogBM_{k,t}$), and cumulative return from the previous 12 months to the previous 2 months as of day t ($Momentum_{k,t}$) in the set of control variables, $X_{k,t}$. As other papers ([39], [88]) present evidence of a relationship between trading volume and future stock returns, we also include trading volume ($volume(\%)_{k,t}$) as a ratio (in percentage) of trading volume relative to total shares outstanding. By including these control variables in

[16] The objective of the regression can also be achieved by using an interaction term with an information indicator variable, which is equal to one if an observation is accompanied by information release. We examine the results with the information dummy variable, and the finding is quantitatively and qualitatively similar to the result with subsamples. The results with the indicator variable are available upon request.

all regressions, we ensure that our findings are not attributable to the relationship between stock returns and other firm characteristics. In the [58] regression with the above specification, the dependent variable has an overlapping period. For example, the daily regression estimation of the next 5-day cumulative returns on the dependent variables explained above has four days of overlap. Due to these overlapping windows in the dependent variables, we apply the Newey-West standard error correction with a lag of 4, 9, and 19 for the regressions of the next 5-, 10-, and 20-day cumulative returns, respectively.

The regression results are provided in Table 32. As stated in the data description in Section 3.2, we use separate subsamples, the *Information sample* and the *No-information sample*, for the [58] regressions because we are interested in the impact of information releases on the role of realized moments in explaining subsequent returns. Panel A and Panel B show the results of the [58] regressions on the *Information sample* and the *No-information sample*, respectively. Comparing the coefficient estimates for realized daily skewness ($RDSkew_{k,t}$) and contemporaneous return ($Ret_{k,t}$) yields one of the main findings of the present paper: even after controlling for contemporaneous returns, realized daily skewness computed using high-frequency returns plays a significant role in explaining subsequent stock returns. Furthermore, the statistical significance and economic significance of realized skewness are strengthened when ambiguous information is released, which leads us to regard realized skewness as a proxy for the information ambiguity concerning a firm.

[109] documents that information ambiguity concerning firms (which is measured by firm size, firm age, analyst coverage, analyst forecast dispersion, and return volatility) generates the cross-sectional variation in return continuations. To investigate our finding with respect to the argument of [109], we further separate the *Information sample* and the *No-information sample* into samples with negative skewness and others with positive skewness.

Table 33 reports the regression results for the subdivided samples. Consistent with our previous finding, we document that realized daily skewness plays a significant role in explaining subsequent stock returns in the sample with negative skewness (Panel A), regardless of the presence of ambiguous information releases. The significant and positive coefficients for realized daily skewness in Panel A confirm the finding of [109]. In Panel B, we expect and find insignificant estimates because the observations in Panel B do not exhibit a strong under-reaction by investors in the absence of greater information ambiguity, represented by positive realized skewness. While the larger magnitude of coefficients in the *Information sample* relative to the *No-information sample* is expected, we also observe significant positive coefficients in the *No-information sample* in Panel A, which indicates that the significance of realized daily skewness as a measure of information ambiguity is not limited to days with information releases under our consideration.

While [109] finds that greater information ambiguity generates greater price drift, [51] predict that ambiguity-averse investors require compensation for information ambiguity in the form of higher expected returns. Our finding supports that of [109] that greater information ambiguity, measured by realized skewness, predicts lower subsequent returns by generating stronger under-reaction by investors. To the best of our knowledge, the present paper is the only one to document that realized daily skewness indeed captures the information ambiguity concerning a firm and further relates it to the cross-sectional variation in return continuations on information days. A detailed analysis of realized skewness as a proxy for information ambiguity is provided in Section 3.5.

3.4 Economic Significance of Main Findings

In previous sections, we show that the realized daily skewness, which captures information ambiguity about a firm's fundamentals, positively predicts subsequent stock returns in the presence of intangible information releases. In addition, we document that return reversals prevail on no-information days, whereas the positive relationship between realized skewness and subsequent returns around information days is robust to the inclusion of contemporaneous return. In this section, we design a zero-net investment trading strategy based on the main finding of the present paper and demonstrate the profitability of the strategy.

Our main finding is on a daily basis, as the role of realized skewness as a proxy for information ambiguity is driven largely by information releases, which should be considered at a high-frequency level to precisely capture the impact. However, employing a daily trading strategy to incorporate our finding would entail tremendous transaction costs. Thus, we propose a monthly trading strategy that utilizes a role of realized skewness in the *Information sample* and a role of contemporaneous returns in the *No-information sample*.

Specifically, over the last week of each month, we first examine whether firms experience earnings forecasts or analyst recommendations releases. For firms with these information releases, we further examine whether their average realized skewness over the week is positive or negative to exploit the predictive ability of realized skewness on information days. For firms without that information, we instead examine whether their average return over the same week is positive or negative to exploit the strong return reversals on no-information days. Based on these two criteria, we construct four portfolios at the end of each month: no-information with positive previous-week return, no-information with negative previous-week return, information with positive skewness, and information with negative skewness portfolios.

Using these four portfolios, we construct two zero-investment portfolios and hold them over the next month: *No-information portfolio* (which takes a long position on the no-information with a negative previous-week return portfolio and a short position on the no-information with a positive previous-week return portfolio) and

Information portfolio (which takes a long position on the information with a positive previous-week skewness portfolio and a short position on the information with a negative previous-week skewness portfolio). As these two portfolios separately utilize our finding on the negative and the positive relationships, we construct another portfolio, called *Combined portfolio*, which takes long positions on the no-information with negative previous-week return and the information with positive previous-week skewness portfolios and short positions on the no-information with positive previous-week return and the information with negative previous-week skewness portfolios.[17] We expect to obtain positive returns from both *No-information portfolio* and *Information portfolio* because of the return reversals and the role of skewness as a proxy for information ambiguity, respectively. In addition to these two zero-investment portfolios, *Combined portfolio* exploits both simultaneously; thus, it is expected to yield a higher return than either *No-information portfolio* or *Information portfolio.*

Table 34 provides summary statistics and pairwise Pearson correlation coefficients of returns on the three zero-investment portfolios (*No-Info*, *Info*, and *Combined*) and other well-known zero-investment risk factors, including MKT (market portfolio in excess a risk-free rate), SMB (small-minus-big), HML (high-minus-low), and UMD (up-minus-down). *Information portfolio* (*No-information portfolio*) yields a 0.20% (0.63%) monthly return with a 1.35% (1.94%) monthly standard deviation. While the monthly returns of *No-information* and *Information* portfolios are not remarkably high relative to other well-known risk factors, the Sharpe ratios (annualized) of these two portfolios are higher than those of other portfolios due to their substantially low volatility. The results for *Combined portfolio* are even more interesting. *Combined portfolio* provides a 0.83% monthly return with a 1.62% standard deviation, resulting in the highest Sharpe ratio, 1.766, of all zero-investment portfolios.

[17]The three zero-investment portfolios, *Information*, *No-information*, and *Combined*, are constructed with equal weights. Our finding is robust to value-weighted portfolios.

Figure 4 illustrates the performance of the zero-investment trading strategies by plotting the cumulative returns on the portfolios beginning from January 2001 to June 2014.[18] It is clear that the three zero-investment strategies experience much less volatile paths relative to other portfolios over the sample period, even during the recent financial crisis. An investment of $1 in *Information portfolio* (*No-information portfolio*) over 161 months yields $1.35 ($2.67) at the end of June 2014, while investing the same amount in *Combined portfolio* yields $3.69 over the same horizon. The time-series plot in Figure 4 displays the superior performance of *Combined portfolio* (solid red line), and its remarkable performance during the recent financial crisis is even more noteworthy.

From Figure 4, it appears that most of the profitability of *Combined portfolio* is attributable to a return reversal trading strategy (*No-information portfolio*). However, the stable performance of *Information portfolio* makes a significant contribution to generating the higher Sharpe ratio of *Combined portfolio*. A strong negative correlation coefficient (-0.567) between *Information portfolio* and *No-information portfolio* indicates that our finding on realized skewness largely complements *Combined portfolio* with higher monthly returns and lower volatility, resulting in an approximately 57% increase in the Sharpe ratio, from 1.126 to 1.766.

In Table 35, we examine whether the performance of the three zero-investment portfolios is due to compensation for the existing risk factors. To test this, we employ the [58] regressions of zero-investment portfolio returns on the common risk factors (MKT, SMB, HML, and UMD). Panel A, Panel B, and Panel C of Table 35 report raw returns and alphas from the [56] three-factor model and from the [30] 4-factor model of *No-information*, *Information*, and *Combined portfolios*. Both returns from

[18] The last period of portfolio construction is the end of May 2014. Thus, the analysis on portfolio performances ends at the end of June 2014.

No-information and *Information portfolios* only have significant exposure to the market factor (MKT, market excess return), while being neutral to other systematic risk factors. More interestingly, the return on *Combined portfolio*, which yields the best performance, does not have any significant exposure to the common risk factors. Thus, consistent with the findings in Figure 4 and Table 34, the performance of *Combined portfolio* is independent of well-known systematic risk factors and is superior throughout the sample period.

3.5 Realized Skewness as a Proxy for Information Ambiguity

In this section, we examine how realized skewness is linked to other existing alternative measures of information ambiguity and show that realized skewness captures a unique feature of information ambiguity beyond existing proxies.

The finding that realized skewness has a strong predictive power in explaining subsequent stock returns on information days suggests that realized skewness captures information ambiguity concerning a firm even at a daily frequency. To test this implication, we collect alternative measures of information ambiguity in the existing literature and examine their relationship with realized skewness. As discussed previously, there are few measures of information ambiguity in the literature with higher than monthly frequencies and with a theoretical guidance and directional information. For example, [109] uses firm size, firm age, analyst forecast dispersion, analyst coverage, stock return volatility, and cash flow volatility as proxies for information ambiguity.

A release of information contains a signal of the true value of a firm with noise. The notion of information ambiguity in this paper is in regard to the difficulty of interpreting the signal, which does not distinguish the true value of the signal from the noise component. As the literature on information asymmetry focuses on information ambiguity regarding a firm's fundamentals, it is inevitable that realized skewness shares some common features of existing proxies for information asymmetry, that are also available at frequencies comparable to our study. Thus, we borrow existing proxies from the literature on information asymmetry. We implicitly assume that these proxies measure the ambiguity of information because information asymmetry is directly related to the quality of information.

In the literature on information asymmetry, [63] show that the bid-ask spread contains a component related to asymmetric information. [106] document that asymmetric information, measured by the bid-ask spread, increases around information releases. Regarding liquidity and trading volume, [26] document that asymmetric information between informed and uninformed investors generates illiquidity. [46] show that the probability of informed trading is negatively related to the trading volume of a stock. Following the finding of [10] on the idiosyncratic volatility puzzle that stocks with high idiosyncratic volatility yield low future returns, [84] further document that stocks with high idiosyncratic volatility predict low future earnings due to uncertainty from a firm's information disclosure. Based on existing findings in the literature, we compare our proposed measure of information ambiguity – realized skewness – with the bid-ask spread (*BAspread*), idiosyncratic volatility of stock returns (*Idio vol*), [5] illiquidity measure (*Amihud*), and change in the turnover ratio (Δ*Turnover*) as existing proxies for information asymmetry.[19]

As a first step to confirm whether a measure of realized daily skewness can serve as a proxy for information ambiguity, we examine the relationship between realized daily skewness and existing proxies for information ambiguity (asymmetry).[20] Table 36 reports regression coefficients for realized daily skewness on the bid-ask spread (*BAspread*), idiosyncratic volatility (*Idio vol*), Amihud liquidity measure (*Amihud*),

[19]Other alternative measures of information ambiguity or divergence of opinion are analyst coverage and analyst forecast dispersion. Due to sample limitations and because these measures are at a lower-than-daily frequency, we are unable to use those measures in our analysis.

[20]We confirm that these existing proxies for information asymmetry play some role in explaining subsequent returns around information releases. The results are available upon request.

and change in turnover ratio ($\Delta Turnover$). From the results in Table 36, we argue that realized skewness can be a proxy for information ambiguity, but we need to interpret the result with caution. Although these existing proxies are well documented to be strongly related to information asymmetry and ambiguity concerning a firm, the direction of the relationship between information ambiguity and these proxies is unclear. While it is unambiguous that the information releases considered in this paper increase the ambiguity of a firm's value, there is no conclusive evidence regarding how these existing proxies should behave in this context.

The interpretation of the bid-ask spread, idiosyncratic volatility, and the illiquidity measure is straightforward. The increase in information ambiguity widens the bid-ask spread and increases idiosyncratic volatility and illiquidity. As [51]'s model predicts, realized daily skewness should decrease as information ambiguity increases. The significant negative regression coefficients on *BAspread*, *Idio vol*, and *Amihud* confirm this prediction. Because the direction of realized daily skewness given an increase in information ambiguity is clear, we are able to further infer how other existing proxies related to information ambiguity actually behave. The negative and significant coefficients on the interaction terms between these three proxies and an information indicator variable, which is equal to one when there is an information release, further confirm this relationship.

The relationship between realized daily skewness and changes in the turnover ratio requires further attention. On normal days, there is a significantly positive relationship between changes in turnover and realized daily skewness, suggesting that an increase in changes in turnover indicates a decrease in information ambiguity. However, when there are large inflows of ambiguous information regarding a firm, this positive relationship becomes negative ($\Delta Turnvoer \times d\_Info$).[21] [31] documents that

[21] [36] investigate the determinants of negative skewness in stock returns and find that large differences in investors' opinions predict (measured by the increase in trading volume) negative skewness in the next six months. While their main objective is to forecast a substantial decline in stock returns

the trading volume of a stock is negatively correlated with information asymmetry before scheduled announcements and positively correlated with asymmetry after such announcements. Thus, our finding regarding the relationship between realized daily skewness and changes in turnover and on the differential shape of this relationship in the context of ambiguous information releases further confirms the finding of [31]. Overall, the findings in Table 36 provide evidence that realized daily skewness, as a proxy for information ambiguity, captures the common features of information ambiguity and information asymmetry, similar to existing proxies.

Realized daily skewness computed with high-frequency returns has a unique and complementary feature that other existing proxies for information ambiguity or asymmetry do not provide. Because high-frequency returns are used to construct realized daily skewness, it contains the intra-day dynamics of return series and information on the distribution of returns, whereas the bid-ask spread changes as a dealer (or a market maker) updates his belief concerning the expected gains from uninformed traders and the expected losses to informed traders. That is, when a dealer expects greater losses to informed traders due to an influx of information, he will increase the bid-ask spread to offset such losses. Thus, the magnitude of the bid-ask spread is directly related to the degree of information asymmetry (see [40] and [64]).

The measures related to liquidity, the [5] illiquidity measure and turnover ratio, are supposed to capture the ease of trading a given stock. Liquidity and turnover are directly associated with the trading dynamics of a dealer or a broker, especially around informational events. Furthermore, idiosyncratic volatility is proven to capture information asymmetry of a firm in the context of asset pricing and the corporate finance literature because higher idiosyncratic volatility implies greater difficulties in

measured by negative skewness, our main interest is in the role of realized skewness as a measure of information ambiguity at a daily frequency. Thus, the relationship between realized skewness and the turnover ratio in this section is contemporaneous in an effort to observe the connection among measures of information ambiguity, whereas the aim of their specification is to predict negative skewness using trading volume at a semi-annual frequency.

evaluating the value of a firm (see, for example, [37] and [93]). Compared to these existing proxies, our proposed measure of information ambiguity, realized daily skewness, enables us to capture richer information on the intra-day returns distribution. The shape of the returns distribution contains valuable information on the divergence of opinion among investors and investors' expectations regarding tail events. Thus, we argue that realized daily skewness makes it possible to capture exclusive components of information ambiguity regarding a firm.

To demonstrate any complementary features of realized daily skewness as a proxy for information ambiguity over existing measures, we take a projection of realized daily skewness on four other alternative proxies borrowed from the literature on information. That is, we extract an orthogonal component of realized daily skewness, called *OrthRDSkew*, that accounts for the effect of existing measures. If realized skewness has a unique contribution in measuring the ambiguity of information concerning a firm beyond other existing proxies, a coefficient estimate on *OrthRDSkew* in predicting subsequent stock returns should remain statistically significant. Table 37 shows the [58] regression result of subsequent stock returns on orthogonalized realized skewness and other control variables. When there is no information release, the significance of realized skewness remains over a short time period ($Ret_{k,t+1,t+5}$). The coefficient estimates are stronger and significant over 10 days on information days, which is consistent with our main finding. Thus, the results in Table 36 and 37 suggest that realized daily skewness makes a significant contribution to explaining the cross-sectional variation in investors' under-reaction as a proxy for information ambiguity beyond other alternative proxies for information ambiguity.

3.6 Robustness Checks

3.6.1 Drift-adjusted Realized Higher Moments

The implicit assumption in computing realized daily moments using equations (25), (26), and (27) is that the high-frequency return converges to zero as the sampling frequency increases. If this assumption is violated because of discrete sampling in the implementation, the findings on realized skewness as a proxy for information ambiguity and its role in explaining return continuations might be contaminated. To mitigate this concern, we compute drift-adjusted realized moments following [4] and repeat the analysis in Section 3.3 with drift-adjusted measures. Drift-adjusted realized daily skewness, variance, volatility, and kurtosis are computed as follows

$$AdjRDSkew_{k,t} = \frac{\sqrt{n}\sum_{l=2}^{n}\left(r_{k,t,l} - \frac{\mu_{k,t}}{n}\right)^3}{AdjRDVar_{k,t}^{3/2}}, \tag{29}$$

$$AdjRDVar_{k,t} = \sum_{l=1}^{n}\left(r_{k,t,l} - \frac{\mu_{k,t}}{n}\right)^2, \tag{30}$$

$$AdjRDVol_{k,t} = \sqrt{AdjRDVar_{k,t}}, \tag{31}$$

$$AdjRDKurt_{k,t} = \frac{n\sum_{l=2}^{n}\left(r_{k,t,l} - \frac{\mu_{k,t}}{n}\right)^4}{AdjRDVar_{k,t}^{2}}, \tag{32}$$

where $\mu_{k,t}$ is the daily return of firm k on day t. That is, the mean-adjusted realized volatility, realized skewness, and realized kurtosis are computed using drift-adjusted 5-minute returns.

Table 38 reports the coefficient estimates from the [58] regression of equation (28) using drift-adjusted realized skewness (equation (29)) instead of realized daily skewness (equation (25)). Comparing the results in Panel B of Table 32, we still observe statistical significance for $AdjRDSkew_{k,t}$, and the significance and magnitude are stronger in the *Information sample.* Despite the use of a drift-adjusted measure of realized skewness, the positive relationship with subsequent returns becomes stronger than the result with the original measure of realized skewness. The findings in Table

38 further confirm that the impact of realized skewness on information days is not merely the return continuation of daily returns.

3.6.2 Reversals/Continuations of Major Price Changes

In a recent contribution, [102] documents that large price changes accompanied by information (using analyst reports as a proxy) are followed by drift (momentum), while those unaccompanied by information experience reversals. Other papers, including [25] and [41], also examine price dynamics following large price changes. To compare our finding with the results in the existing literature, we conduct our main analysis after excluding large price changes. We filter out observations if their daily returns are greater than 10% (5%) or less than -10% (-5%). These ±10% and ±5% filtering requirements for large price changes exclude 79,054 and 433,995 observations, which represent 1.31% and 7.17% of the total number of observations, respectively.

In Panel A (Panel B) of Table 39, we examine our main results for the samples without ±10% (±5%) changes in daily stock prices. The result also includes the analysis with drift-adjusted realized skewness *AdjRDSKew*. From the results in Table 39, it is evident that realized skewness plays a significant role in predicting subsequent returns, and its role becomes further pronounced on information days, even after excluding large stock price changes. Regarding the results with drift-adjusted skewness and daily return, the positive relationship strongly persists even after controlling for the return continuations, whereas the negative relationship appears to be attributable to return reversals.

3.6.3 Liquidity Provision of Market Makers

Liquidity has been documented as an important attribute of stock returns (see, for example, [6], among many others). Around the news release, informed traders have a superior informational advantage to trade even before the news releases. Due to this information asymmetry, market makers would be wary of adverse price changes around the news events. Therefore, market makers would not only demand high compensation for adverse price changes but also be reluctant to provide liquidity before the events. From this evidence in the literature on liquidity and market microstructure, one might doubt that our finding on realized skewness as a proxy for information ambiguity is manifestation of liquidity concerns of market makers around information events.

[94] shows that the liquidity provision of market makers can be well proxied using lagged stock returns. Thus, to test the robustness of our finding, we follow [94] and use lagged returns as a noisy proxy for the liquidity provision. Table 40 provides the [58] regression result that includes a lagged return ($Ret_{k,t-3,t-1}$) as an additional control variable in the specification. Consistent with the finding in [94], the liquidity provision, proxied by the lagged return, plays a significant role in explaining the return reversals in the *No-information sample*. However, the statistical significance of realized skewness remains intact in the *Information-sample* as well as in the *No-information sample*. Thus, our finding on realized skewness cannot be explained by the liquidity provision of market makers around the news releases.

3.7 Conclusion

In this paper, we examine the role of realized skewness as a proxy for information ambiguity. We find that realized daily skewness plays an important role in explaining subsequent stock returns, even after controlling for return reversals and continuations. The role of realized skewness becomes strong on information days, and the statistically significant impact of realized skewness on subsequent returns is attributable primarily to observations with negative skewness. We show that realized skewness measures information ambiguity concerning a firm's fundamentals and captures a unique feature of information ambiguity not captured by other alternative proxies. A zero-net investment trading strategy that exploits our main finding yields a superior performance with a Sharpe ratio of 1.766 compared to other well-known zero-investment portfolios over the sample period from January 2001 to June 2014.

The present paper contributes to the literature in several ways. First, it provides evidence that realized daily skewness can be a proxy for information ambiguity about a firm's fundamentals. Although information ambiguity generates significant frictions in relevant markets, there are few proxies with a clear theoretical guidance to measure information ambiguity concerning a firm's fundamentals at higher frequency due to data limitations. We demonstrate that realized daily skewness, which is computed using high-frequency data, is able to assess the ambiguity of information.

Second, we contribute to the literature on return predictability by demonstrating that realized daily skewness, as a measure of information ambiguity, predicts subsequent stock returns. While [109] finds that greater information ambiguity generates greater price drifts, [51] predict that ambiguity-averse investors require compensation for information ambiguity in the form of higher expected returns. Our finding supports that of [109] that greater information ambiguity measured by realized skewness computed with high-frequency data predicts lower subsequent returns by generating stronger under-reaction by investors.

[2] proposes a theoretical asset pricing model to explain contrasting evidence on skewness in stock returns: positive skewness in firm-level stock returns and negative skewness in aggregate stock market returns. He argues that the cross-sectional heterogeneity in the timing of firms' cash-flow news leads to negative skewness in aggregate stock returns, while firm-level stock returns exhibit positive skewness. Our finding on realized skewness as a proxy for information ambiguity might also be extended to explain the difference in realized skewness between firm-level and aggregate-level stock returns. We might expect to observe greater negative skewness in aggregate stock market returns relative to firm-level stock returns because an aggregate stock

portfolio, such as the S&P 500 index, contains considerably more intangible information (relative to tangible information, for example, cash-flow news) and has a complex information environment.

The findings in this paper shed light on the importance of realized daily moments, such as realized daily volatility, skewness, and kurtosis. Compared to other well-known important asset-pricing variables, such as daily returns and trading volumes, higher-frequency returns are able to provide much richer information on the return distribution of underlying assets than daily or monthly returns. The present paper documents an aspect of realized skewness on the information ambiguity of a firm's fundamentals. Our finding suggests that other higher moments may contain valuable information about a firm, not only regarding information ambiguity but also concerning the trading behavior of informed and uninformed traders, which is left for future research.

Figure 4: Performance of zero-investment trading strategies

The figure presents the performance of zero-investment trading strategies using cumulative gains from January 2001 to June 2014. No-information, Information, and Combined portfolios are the zero-investment portfolios based on the findings of the present paper. The three zero-investment portfolios, *No-information portfolio (Noinfo)*, *Information portfolio (Info)*, and *Combined portfolio (Combined)* are constructed as follows. Over the last week of each month, we first examine whether firms experience analyst earnings forecast or recommendation report releases. In a set of stocks with these information releases, we examine whether their average realized daily skewness over the corresponding week is positive or negative. In the other set of stocks without information releases, we examine whether their average daily returns over the week are positive or negative. Based on these two criteria, we construct four portfolios at the end of every month: no-information with positive skewness, no-information with negative skewness, information with positive skewness, and information with negative skewness portfolios. *Noinfo* is the return on *No-information portfolio* over the next month, which takes a long position on the no-information with negative previous-week return portfolio and a short position on the no-information with positive previous-week return portfolio. *Info* is the return on *Information portfolio* over the next month, which takes a long position on the information with positive previous-week skewness portfolio and a short position on the information with negative previous-week skewness portfolio. *Combined* is the return on *Combined portfolio* constructed by taking a long position on *No-information portfolio* and a short position on *Information portfolio*.The plots of *MKT*, *SMB*, *HML*, and *UMD* represent the cumulative gains from the market excess return, small-minus-big, high-minus-low, and up-minus-down portfolios (see [56] and [30]).

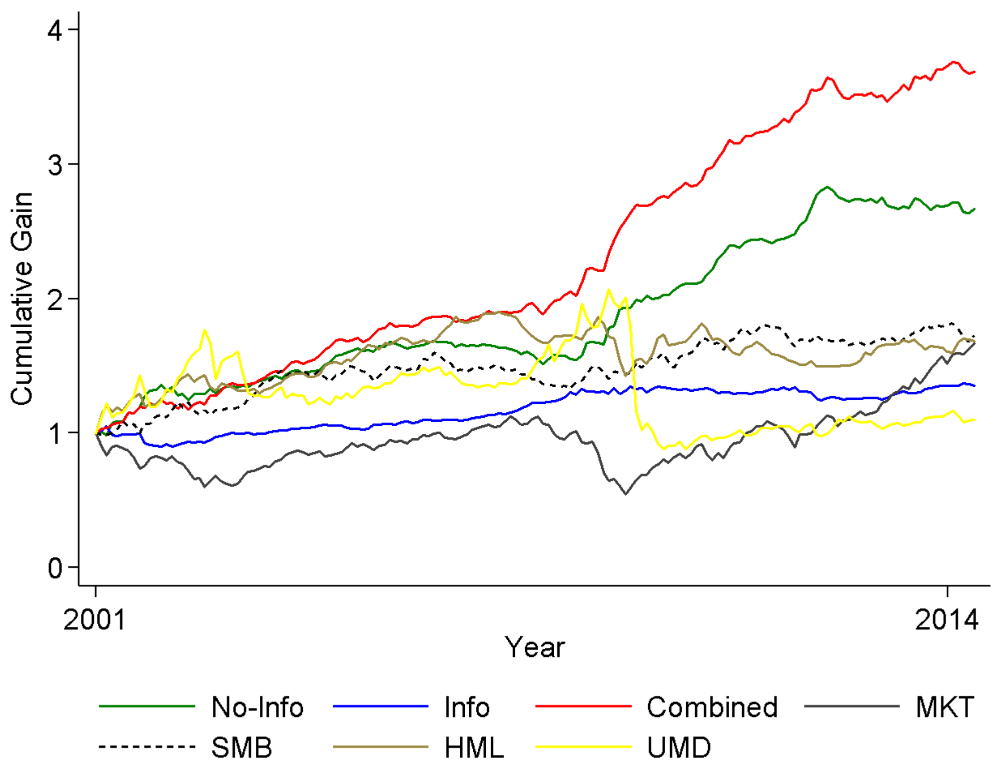


Table 28: Summary statistics of the sample

This table provides time-series summary statistics (means and medians) of variables for the full sample (from January 2001 to May 2014) used in this paper. The first column under *Firms* reports the number of firms in the sample in each year. The next two columns, *Daily return*, presents the average daily returns of stocks computed from the CRSP daily file. *Realized Vol*, *Realized Skew*, and *Realized Kurt* are realized daily volatility (equation (26)), realized daily skewness (equation (25)), and realized daily kurtosis (equation (27)) computed using 5-minute returns from the TAQ database, respectively. *Market equity* represents the average market capitalization calculated using the closing market prices of firms. *Book-to-Market* is the ratio of book equity to market equity of firms. The variable *Volume (%)* is computed as the percentage of daily trading volume relative to total shares outstanding. In 5-minute returns and higher moment calculations, stocks with prices lower than $5 are excluded. To be included in the sample, a stock must have at least 100 transactions within a trading day.

	Firms	*Daily return*		*Realized Vol*		*Realized Skew*		*Realized Kurt*		*Market equity*		*Book-to-Market*		*Volume (%)*	
Year		Mean	Med	Mean	Med	Mean	Med	Mean	Med	Mean	Med	Mean	Med	Mean	Med
2001	3391	0.22%	0.02%	0.038	0.030	0.021	0.028	8.22	6.62	5994.4	1160.7	1.409	0.376	0.011	0.006
2002	3251	-0.03%	-0.10%	0.032	0.026	-0.026	-0.018	7.89	6.31	4963.9	1013.9	1.722	0.461	0.010	0.005
2003	3474	0.27%	0.12%	0.024	0.019	0.012	0.014	7.96	6.29	4744.4	993.3	1.609	0.468	0.010	0.006
2004	3716	0.16%	0.07%	0.022	0.017	0.050	0.046	7.88	6.15	5140.1	1094.9	1.255	0.402	0.011	0.006
2005	3903	0.11%	0.02%	0.020	0.017	0.010	0.008	7.75	6.01	5276.6	1135.0	1.086	0.388	0.010	0.006
2006	3876	0.09%	0.02%	0.019	0.016	0.030	0.023	7.13	5.51	6285.5	1410.4	1.217	0.396	0.010	0.006
2007	2010	0.03%	0.02%	0.019	0.016	0.011	-0.012	7.49	5.45	8763.8	2227.8	1.468	0.435	0.011	0.007
2008	1906	-0.09%	-0.14%	0.034	0.028	-0.015	-0.007	7.25	5.21	7696.4	1904.5	2.021	0.598	0.015	0.010
2009	1748	0.24%	0.14%	0.028	0.025	-0.014	-0.015	7.22	5.40	6538.0	1705.3	2.886	0.732	0.014	0.009
2010	2341	0.12%	0.08%	0.019	0.016	0.012	-0.007	7.04	5.39	7798.0	2122.1	2.314	0.577	0.012	0.008
2011	2411	0.00%	0.02%	0.020	0.017	0.009	-0.014	6.68	5.13	8731.0	2449.9	2.654	0.539	0.013	0.008
2012	2161	0.10%	0.05%	0.016	0.014	0.029	0.010	7.15	5.52	9226.8	2485.3	4.361	0.581	0.011	0.007
2013	2081	0.13%	0.11%	0.015	0.013	0.040	0.022	7.34	5.69	10556.3	2931.7	3.283	0.500	0.011	0.006
2014	1852	0.05%	0.06%	0.016	0.013	0.017	0.000	7.26	5.62	11294.7	3138.1	2.845	0.456	0.012	0.007
Total		0.10%	0.04%	0.023	0.018	0.014	0.007	7.48	5.77	6981.9	1594.3	2.019	0.470	0.011	0.007

Table 29: Time-series distribution of firm-specific information releases

This table describes our firm-day observations with and without firm-specific information releases over our sample period from January 2001 to May 2014. Two columns under *Full sample* list the number of all firm-day observations and the average size of firms included in the sample in each year. The next four columns provide information about the subsamples that include firm-day observations accompanied by two different types of firm-specific information releases. The two columns under *Earnings forecast* show the number of firm-day observations accompanied by analyst earnings forecasts (from the I/B/E/S database) and the average size of the corresponding firms. The next two columns under *Analyst recommendation* show the number of firm-day observations accompanied by one or more analyst recommendation reports and the average size of the corresponding firms (from the I/B/E/S database). Two columns under *Information* show the number of observations accompanied by either earnings forecasts or analyst recommendations and the average size of those firms. The last two columns under *No-information* provide information about the subsample that includes firm-day observations that are not accompanied by either of the two types of firm-specific information releases we consider in this study (i.e., earnings forecasts and analyst recommendations).

	Full sample		*Earnings forecast*		*Analyst recommendation*		*Information*		*No-information*	
Year	Obs	Size	Obs	Size	Obs	Size	Obs	Size	Obs	Size
2001	408,718	5,994.37	88,550	10,529.42	45,036	10,197.69	105,973	10,129.58	302,745	4,546.87
2002	461,887	4,963.91	92,863	8,476.88	68,565	8,979.40	127,033	8,316.79	334,854	3,691.94
2003	503,585	4,744.41	105,780	8,532.14	60,619	8,658.30	128,047	8,306.05	375,538	3,530.00
2004	569,511	5,140.07	117,505	8,960.11	58,574	8,997.16	137,444	8,805.03	432,067	3,974.22
2005	627,514	5,276.56	126,287	9,611.27	55,769	8,785.83	144,222	9,306.31	483,292	4,074.01
2006	542,254	6,285.50	110,489	11,117.43	48,168	10,560.50	128,016	10,878.37	414,238	4,866.12
2007	404,243	8,763.75	88,926	14,559.12	38,500	13,529.01	103,708	14,163.94	300,535	6,900.27
2008	387,852	7,696.45	103,045	12,403.80	43,992	12,451.87	117,992	12,211.12	269,860	5,722.48
2009	351,856	6,538.02	98,539	10,188.19	39,460	9,916.40	109,385	9,983.96	242,471	4,983.47
2010	401,804	7,798.01	112,912	12,025.92	43,059	12,159.03	124,557	11,901.08	277,247	5,954.65
2011	435,201	8,730.97	127,950	13,610.91	49,268	13,172.88	140,886	13,309.00	294,315	6,539.50
2012	394,969	9,226.84	116,400	14,239.51	40,923	13,820.16	127,604	13,915.36	267,365	6,989.18
2013	401,784	10,556.33	112,780	16,238.82	36,973	15,637.33	122,818	15,929.12	278,966	8,190.90
2014	163,688	11,294.66	47,961	17,553.08	14,861	17,260.09	51,966	17,262.05	111,722	8,519.01
Total	6,054,866	7,357.85	1,449,987	12,003.33	643,767	11,723.26	1,669,651	11,744.13	4,385,215	5,605.90

Table 30: Impact of information on realized daily moments

This table shows the impact of information release on daily return and realized daily moments. *Realized daily skewness*, *realized daily volatility*, and *realized daily kurtosis* are computed using 5-minute returns and equations (25), (26), and (27). The column under *Full* reports the averages of realized daily moments. The next two columns, *Forecast* and *Recommendation*, show the averages when the observations are accompanied by earnings forecasts and by analyst recommendations, respectively. *Information* and *No-information* report the averages of moments of the *Information* sample and those of the *No-information* sample. The last column reports the differences in moments between values reported in *Information* and *No-information*. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Realized moments in each subsample

	Full	*Forecast*	*Recommendation*	*Information*	*No-information*	*Diff(Info - No-info)*
RDSkew	0.0144	0.0073	0.0075	0.0076	0.0170	-0.0094***
Returns	0.0010	0.0007	0.0009	0.0009	0.0011	-0.0002***
RDVol	0.0229	0.0235	0.0251	0.0236	0.0227	0.0009***
RDKurt	7.4831	7.3346	7.6031	7.3697	7.5263	-0.1565***

Panel B. Pairwise correlation coefficients

	Daily Return	*Realized Vol*	*Realized Skew*
Realized Vol	0.024		
Realized Skew	0.408	-0.009	
Realized Kurt	0.009	0.198	0.007

Table 31: Realized skewness and future returns: Controlling contemporaneous returns

This table reports next 5-, 10-, and 20-day portfolio returns, where portfolios are formed based on sorts into three terciles of daily returns (Ret_t). Within each of these terciles, all firms are sorted into terciles of realized daily skewness ($RDSkew_t$). All portfolios are constructed with equal weights. Panel A, Panel B, and Panel C report the double-sorted portfolio returns for the *Full sample*, the *Information sample*, and the *No-information sample*, respectively. Stocks with a price less than five dollars are excluded. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Full sample

	Ret_t			
Next 5-day return				
$RDSkew_t$	1 (Low)		3 (High)	H-L
1 (Low)	46.81	30.54	25.20	-21.61***
				(-7.13)
2	45.13	29.58	22.74	-22.39***
				(-7.19)
3 (High)	47.73	32.05	27.95	-19.78***
				(-6.52)
H-L	0.92	1.50	2.75*	
t-stat	(0.58)	(1.19)	(1.83)	
Next 10-day return				
$RDSkew_t$	1 (Low)		3 (High)	H-L
1 (Low)	77.45	55.34	51.65	-25.80***
				(-6.62)
2	74.68	52.99	47.08	-27.60***
				(-6.98)
3 (High)	78.77	57.81	55.34	-23.42***
				(-6.06)
H-L	1.32	2.47	3.69*	
t-stat	(0.62)	(1.47)	(1.88)	
Next 20-day return				
$RDSkew_t$	1 (Low)		3 (High)	H-L
1 (Low)	139.01	108.30	105.78	-33.24***
				(-6.45)
2	132.42	105.92	99.36	-33.06***
				(-6.15)
3 (High)	139.23	111.41	110.81	-28.42***
				(-5.55)
H-L	0.22	3.11	5.04*	
t-stat	(0.08)	(1.34)	(1.90)	

Panel B. Information sample

	Ret_t			
Next 5-day return				
$RDSkew_t$	1 (Low)		3 (High)	H-L
1 (Low)	32.65	30.12	27.38	-5.27
				(-1.42)
2	36.37	29.64	25.81	-10.56***
				(-2.70)
3 (High)	40.68	32.25	34.48	-6.20*
				(-1.68)
H-L	8.03***	2.13	7.10***	
t-stat	(3.34)	(1.05)	(2.94)	
Next 10-day return				
$RDSkew_t$	1 (Low)		3 (High)	H-L
1 (Low)	58.60	54.46	53.91	-4.69
				(-0.96)
2	62.83	53.41	48.14	-14.69***
				(-2.91)
3 (High)	70.54	57.00	63.84	-6.70
				(-1.40)
H-L	11.94***	2.54	9.93***	-2.01
t-stat	(3.69)	(0.96)	(3.19)	(-0.46)
Next 20-day return				
$RDSkew_t$	1 (Low)		3 (High)	H-L
1 (Low)	111.22	102.47	105.89	-5.33
				(-0.83)
2	114.94	103.57	98.60	-16.34**
				(-2.40)
3 (High)	121.81	107.07	116.58	-5.23
				(-0.82)
H-L	10.60**	4.60	10.69**	0.10
t-stat	(2.44)	(1.29)	(2.48)	(0.02)

Panel C. No-information sample

	Ret_t			
Next 5-day return				
$RDSkew_t$	1 (Low)		3 (High)	H-L
1 (Low)	51.99	30.66	23.95	-28.05***
				(-9.23)
2	48.10	29.19	21.29	-26.80***
				(-8.74)
3 (High)	50.39	31.50	25.08	-25.31***
				(-8.24)
H-L	-1.60	0.84	1.14	2.74
t-stat	(-0.93)	(0.64)	(0.71)	(1.19)
Next 10-day return				
$RDSkew_t$	1 (Low)		3 (High)	H-L
1 (Low)	84.60	55.87	50.50	-34.10***
				(-8.75)
2	78.21	51.87	45.25	-32.96***
				(-8.44)
3 (High)	81.54	57.83	52.17	-29.37***
				(-7.52)
H-L	-3.06	1.96	1.67	4.74
t-stat	(-1.32)	(1.08)	(0.79)	(1.55)
Next 20-day return				
$RDSkew_t$	1 (Low)		3 (High)	H-L
1 (Low)	150.05	110.57	106.35	-43.70***
				(-8.38)
2	139.01	104.66	98.34	-40.67***
				(-7.64)
3 (High)	144.77	112.87	108.42	-36.35***
				(-6.98)
H-L	-5.28*	2.30	2.07	7.35*
t-stat	(-1.70)	(0.89)	(0.72)	(1.82)

Table 32: Realized skewness and future stock returns: Information vs. No-information

This table investigates the patterns of relationships between realized daily higher moments and future stock returns depending on the information releases at the firm-day observation level. The dependent variables are the next 5-, 10-, and 20-day cumulative returns ($Ret_{k,t+1,t+5}$, $Ret_{k,t+1,t+10}$, and $Ret_{k,t+1,t+20}$). Realized daily skewness ($RDSkew_{k,t}$), realized daily volatility ($RDVol_{k,t}$), and realized daily kurtosis ($RDKurt_{k,t}$) are computed using 5-minute returns with equations (25), (26), and (27), respectively. A contemporaneous return $Ret_{k,t}$ is included to control for the well-known factors of return reversals and continuations. In addition to daily higher moments, the market size of a firm ($LogME_{k,t}$), the book-to-market ratio ($LogBM_{k,t}$), the cumulative returns over 11 months from the previous 12-month to the previous 2-month period ($Momentum_{k,t}$), and the percentage of daily trading volume ($Volume(\%)_{k,t}$) are included as control variables. The coefficient estimates and t-statistics are computed using the Fama-MacBeth (1973) regression. Because of a possible autocorrelation structure in overlapping windows on dependent variables, we employ the Newey-West standard error correction with corresponding horizon for future stock returns. That is, in the regression with next 5-, 10-, and 20-day returns, we employ the Newey-West correction with a lag of 4, 9, or 19, respectively. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Panel A. Information sample			***Panel B. No-information sample***		
	$Ret_{t+1,t+5}$	$Ret_{t+1,t+10}$	$Ret_{t+1,t+20}$	$Ret_{t+1,t+5}$	$Ret_{t+1,t+10}$	$Ret_{t+1,t+20}$
RDSkew	1.88***	2.51***	1.93*	0.74**	0.52	-0.13
	(3.25)	(3.20)	(1.96)	(2.04)	(1.04)	(-0.17)
RDVol	386.03*	754.29**	1328.43**	676.30***	1156.38***	2155.68***
	(1.82)	(2.11)	(2.07)	(3.31)	(3.34)	(3.33)
RDKurt	-0.38*	-0.56	-1.08*	-0.41**	-0.45	-0.80
	(-1.67)	(-1.55)	(-1.78)	(-2.19)	(-1.46)	(-1.53)
Ret	0.00	0.00	0.01	-0.05***	-0.05***	-0.06***
	(0.39)	(0.50)	(1.40)	(-12.34)	(-8.78)	(-6.77)
LogME	-6.16***	-10.87***	-20.19***	-6.17***	-11.74***	-21.48***
	(-6.09)	(-6.01)	(-5.84)	(-6.72)	(-6.72)	(-6.21)
LogBM	5.02***	7.70***	13.98**	3.46***	6.76***	13.69***
	(3.32)	(2.73)	(2.38)	(2.95)	(2.86)	(2.80)
Momentum	-0.10	-4.78	-13.17	-3.19	-8.59	-16.66
	(-0.01)	(-0.31)	(-0.42)	(-0.55)	(-0.73)	(-0.66)
Volume (%)	-235.92***	-456.95***	-908.38***	29.49	-162.31	-580.76**
	(-2.70)	(-3.35)	(-3.88)	(0.41)	(-1.22)	(-2.47)
Constant	69.18***	122.21***	236.63***	59.78***	114.43***	220.02***
	(5.69)	(5.18)	(4.95)	(5.85)	(5.66)	(5.40)
Observations	1280100	1280100	1280100	4774766	4774766	4774766
R^2	0.10	0.10	0.10	0.07	0.06	0.06

Table 33: Subsample analysis on realized skewness and subsequent returns: Positive and negative skewness samples

This table examines the relationship between realized skewness and subsequent stock returns in separate subsamples: one with negative realized daily skewness (Panel A) and another with positive realized daily skewness (Panel B). The dependent variables are the next 5-, 10-, and 20-day cumulative returns ($Ret_{k,t+1,t+5}$, $Ret_{k,t+1,t+10}$, and $Ret_{k,t+1,t+20}$). Realized daily skewness ($RDSkew_{k,t}$), realized daily volatility ($RDVol_{k,t}$), and realized daily kurtosis ($RDKurt_{k,t}$) are computed using 5-minute returns with equations (25), (26), and (27), respectively. A contemporaneous return $Ret_{k,t}$ is included to control for the well-known factors of return reversals and continuations. In addition to daily higher moments, the market size of a firm ($LogME_{k,t}$), the book-to-market ratio ($LogBM_{k,t}$), the cumulative returns over the 11 months from the previous 12-month to the previous 2-month period ($Momentum_{k,t}$), and the percentage of daily trading volume ($Volume(\%)_{k,t}$) are included as control variables. The coefficient estimates and corresponding t-statistics are computed using Fama-MacBeth (1973) regressions with Newey-West standard errors. In the Newey-West standard error correction, a lag of 4, 9, or 19 is applied to the next 5-, 10-, and 20-day returns, respectively. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Sample with negative skewness

	Information sample			*No-information sample*		
	$Ret_{t+1,t+5}$	$Ret_{t+1,t+10}$	$Ret_{t+1,t+20}$	$Ret_{t+1,t+5}$	$Ret_{t+1,t+10}$	$Ret_{t+1,t+20}$
RDSkew	4.20***	4.74***	6.31***	2.23***	2.36***	2.06*
	(3.76)	(3.30)	(3.18)	(3.37)	(2.71)	(1.79)
Ret	0.00	0.00	0.01	-0.06***	-0.07***	-0.09***
	(0.75)	(0.34)	(1.04)	(-12.84)	(-10.88)	(-10.48)
LogME	-6.14***	-12.31***	-23.35***	-9.01***	-17.03***	-31.49***
	(-7.29)	(-10.98)	(-15.29)	(-12.27)	(-17.10)	(-22.14)
LogBM	3.66***	6.07***	11.24***	2.12***	4.34***	9.61***
	(2.95)	(3.63)	(4.85)	(2.70)	(3.97)	(6.12)
Momentum	0.67	-7.20	-20.65**	-1.68	-9.74*	-20.95***
	(0.12)	(-0.99)	(-1.99)	(-0.45)	(-1.88)	(-2.76)
Volume (%)	-157.80	-355.03**	-724.24***	114.35	-54.00	-309.44**
	(-1.57)	(-2.57)	(-4.11)	(1.58)	(-0.52)	(-2.24)
Constant	77.29***	148.96***	287.49***	91.85***	175.42***	332.10***
	(7.91)	(11.54)	(15.79)	(10.59)	(15.23)	(20.35)
Observations	638129	638129	638129	2371820	2371820	2371820
R^2	0.11	0.11	0.11	0.06	0.06	0.06

Panel B. Sample with positive skewness						
	Information sample			*No-information sample*		
	$Ret_{t+1,t+5}$	$Ret_{t+1,t+10}$	$Ret_{t+1,t+20}$	$Ret_{t+1,t+5}$	$Ret_{t+1,t+10}$	$Ret_{t+1,t+20}$
RDSkew	0.88	2.12	1.95	-0.21	0.48	-1.23
	(0.81)	(1.52)	(1.00)	(-0.34)	(0.57)	(-1.09)
Ret	0.00	0.01	0.01	-0.04***	-0.04***	-0.03***
	(0.68)	(0.81)	(1.11)	(-8.97)	(-6.63)	(-4.24)
LogME	-9.39***	-15.14***	-24.89***	-9.35***	-16.47***	-29.89***
	(-11.11)	(-13.82)	(-14.50)	(-12.80)	(-16.53)	(-20.82)
LogBM	6.28***	8.45***	15.00***	2.12***	4.81***	9.68***
	(5.08)	(4.99)	(6.19)	(2.77)	(4.42)	(6.17)
Momentum	-9.76*	-16.84**	-29.33***	-8.56**	-14.54***	-24.93***
	(-1.79)	(-2.21)	(-2.74)	(-2.25)	(-2.71)	(-3.17)
Volume (%)	-70.73	-263.12**	-489.14***	174.07**	101.27	-224.83*
	(-0.69)	(-1.96)	(-2.75)	(2.38)	(1.02)	(-1.72)
Constant	104.60***	171.90***	296.36***	95.80***	170.14***	320.33***
	(10.83)	(13.57)	(14.98)	(11.11)	(14.78)	(19.43)
Observations	641971	641971	641971	2402946	2402946	2402946
R^2	0.11	0.11	0.11	0.06	0.06	0.06

Table 34: Zero-investment portfolios: Information, No-information, and Combined portfolios

This table reports the summary statistics for the [56] three factors (market excess return (MKT), small-minus-big (SMB), and high-minus-low (HML)), the momentum factor (up-minus-down (UMD)), and the three zero-investment portfolios proposed in the paper over 161 months from February 2001 to June 2014. The three zero-investment portfolios - *No-information portfolio (Noinfo)*, *Information portfolio (Info)*, and *Combined portfolio (Combined)* - are constructed as follows. Over the last week of each month, we first examine whether firms experience earnings forecasts or recommendations releases. For firms with these releases, we further examine whether their average realized skewness over the week is positive or negative. For the firms without that information, we instead examine whether their average return over the same week is positive or negative. Based on these two criteria, we construct four portfolios at the end of each month: no-information with positive return, no-information with negative return, information with positive skewness, and information with negative skewness portfolios. *Noinfo* is the return on the *No-information portfolio* over the next month, which takes a long position on the no-information with negative return portfolio and a short position on the no-information with positive return portfolio. *Info* is the return on the *Information portfolio* over the next month, which takes a long position on the information with positive skewness portfolio and a short position on the information with negative skewness portfolio. *Combined* is the return on the *Combined portfolio* constructed by combining *No-information portfolio* and *Information portfolio* to exploit both the return reversals and the predictability of realized skewness. Panel A reports the summary statistics of each portfolio, and Panel B provides pairwise Pearson correlation coefficients among well-known zero-investment portfolios and the three zero-investment portfolios proposed in the present paper.

Panel A. Summary statistics

Portfolio	Obs	Mean	Std Dev	Minimum	Maximum	Sharpe Ratio
MKT	161	0.42%	4.56%	-17.2%	11.4%	0.321
SMB	161	0.37%	2.53%	-6.6%	6.9%	0.509
HML	161	0.36%	2.72%	-9.9%	13.9%	0.459
UMD	161	0.21%	5.22%	-34.7%	12.5%	0.138
Noinfo	161	0.63%	1.94%	-4.5%	8.8%	1.126
Info	161	0.20%	1.35%	-7.1%	3.3%	0.503
Combined	161	0.83%	1.62%	-3.3%	6.5%	1.766

Panel B. Correlation coefficients

	MKT	SMB	HML	UMD	Noinfo	Info
SMB	0.356					
HML	0.018	0.023				
UMD	-0.500	-0.137	0.041			
Noinfo	0.081	0.009	-0.217	-0.151		
Info	-0.161	-0.172	0.252	0.186	-0.567	
Combined	-0.037	-0.132	-0.050	-0.026	0.727	0.154

Table 35: Calendar-time trading strategy: Zero-investment portfolios

This table presents the profitability of zero-investment portfolios based on our finding of a relationship between realized daily skewness and subsequent returns. The table reports coefficient estimates using Fama-MacBeth (1973) regressions for the following specification: $R^{port}_{t,t+1month} = \alpha^{port} + \beta^{port\prime} X_{t,t+1month} + u^{port}_{t,t+1month}$, where $u^{port}_{t,t+1month}$ is an error term.
The dependent variables are the returns of zero-investment portfolios over one month from the date of portfolio construction (day t), which is the last day of each month. The vector of explanatory variables, $X_{t,t+1month}$, includes common risk factors: *MKT*, *SMB*, *HML*, and *UMD*. The zero-investment trading strategy that is used in this table is as follows. Over the last week of each month, we first examine whether firms experience earnings forecasts or recommendations releases. For firms with these releases, we further examine whether their average realized skewness over the week is positive or negative. For the firms without that information, we instead examine whether their average return over the same week is positive or negative. Based on these two criteria, we construct four portfolios at the end of each month: no-information with positive return, no-information with negative return, information with positive skewness, and information with negative skewness portfolios. Panel A provides the results for the zero-investment strategy that takes a long position on the no-information with negative return portfolio and a short position on the no-information with positive return portfolio, called the *No-information portfolio*. Panel B provides the results of a zero-investment strategy that takes a long position on the information with positive skewness portfolio and a short position on the information with negative skewness portfolio, called the *Information portfolio*. In Panel C, to exploit both the return reversals and the predictability of realized skewness, the *Combined portfolio* is constructed by combining *No-information portfolio* and *Information portfolio*. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

		α	β_{MKT}	β_{SMB}	β_{HML}	β_{UMD}
Panel A. No-information portfolio						
Raw return	Mean	63.0509***				
	t-stat	(4.12)				
CAPM	Mean	61.5962***	0.0344			
	t-stat	(3.91)	(0.80)			
3-Factor	Mean	65.8348***	0.0434	-0.0123	-0.1413	
	t-stat	(4.06)	(1.01)	(-0.15)	(-1.53)	
4-Factor	Mean	67.9585***	0.0111	-0.0040	-0.1405	-0.0539
	t-stat	(4.16)	(0.25)	(-0.05)	(-1.56)	(-1.33)
Panel B. Information portfolio						
Raw return	Mean	19.5961*				
	t-stat	(1.84)				
CAPM	Mean	21.6101**	-0.0476*			
	t-stat	(2.09)	(-1.73)			
3-Factor	Mean	20.3584*	-0.0403	-0.0774	0.1281**	
	t-stat	(-1.96)	(-1.40)	(-1.30)	(2.16)	
4-Factor	Mean	18.8107*	-0.0168	-0.0834	0.1275**	0.0393
	t-stat	(1.75)	(-0.55)	(-1.38)	(2.25)	(1.40)
Panel C. Combined portfolio						
Raw return	Mean	82.6470***				
	t-stat	(6.47)				
CAPM	Mean	83.2063***	-0.0132			
	t-stat	(6.21)	(-0.39)			
3-Factor	Mean	86.1932***	0.0031	-0.0896	-0.0132	
	t-stat	(6.35)	(0.09)	(-1.31)	(-0.21)	
4-Factor	Mean	86.7691***	-0.0056	-0.0874	-0.0130	-0.0146
	t-stat	(6.45)	(-0.15)	(-1.26)	(-0.20)	(-0.49)

Table 36: Realized skewness and information ambiguity

This table examines the relationship between realized daily skewness and various measures of information ambiguity for firms over the period from January 2001 to May 2014. The panel dataset is at the firm-day level. Panel A reports the pairwise correlation coefficients. Panel B reports the panel regression results of realized daily skewness on other proxies for information ambiguity interacting with the information dummy variable, d_Info, which is equal to 1 if a firm-day observation is accompanied by intangible information releases (i.e., an analyst recommendation or earnings forecast) and 0 otherwise. The variable *BAspread* is computed as the ratio of the difference between ask and bid prices to the mid-point of ask and bid prices, and the idiosyncratic volatility of a firm (*Idio vol*) is the standard deviation of residuals of the previous 60-day returns from the [56] three-factor model. *Amihud* is an illiquidity measure of [5]. Δ*Turnover* is the change in the ratio of daily trading volume to the total shares of outstanding. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Correlation matrix

	RDSkew	BAspread	Idio vol	Amihud	ΔTurnover
RDSkew	1.0000				
BAspread	-0.0034	1.0000			
Idio vol	-0.0051	0.2037	1.0000		
Amihud	-0.0116	0.1478	0.5441	1.0000	
ΔTurnover	0.0068	0.0001	-0.0059	0.0232	1.0000

Panel B. Realized daily skewness on information ambiguity

	(1)	(2)	(3)	(4)
BAspread	-0.39***			
	(-3.68)			
BAspread×d_Info	-0.88***			
	(-4.81)			
Idio vol		-0.33***		
		(-8.38)		
Idio vol×d_info		-0.43***		
		(-8.65)		
Amihud			-3.11***	
			(-22.90)	
Amihud×d_info			-1.83***	
			(-11.68)	
ΔTurnover				0.49***
				(20.16)
ΔTurnover×d_info				-0.60***
				(-12.06)
Observations	5897483	6045930	6045949	6039164

Table 37: Incremental role of realized daily skewness: Orthogonalized realized skewness

This table investigates the incremental role of realized daily skewness. To examine the additional contribution of the measure over the alternative proxies for information ambiguity, we orthogonalize realized daily skewness on four daily measures of information asymmetry (*BAspread*, *Idio vol*, *Amihud*, and Δ*Turnover*). That is, the residuals from the [58] regression of realized daily skewness on four proxies are used as orthogonalized realized skewness ($OrthRDSkew_{k,t}$). Using these orthogonalized realized skewness results as the main independent variables, we estimate similar regressions to Table 32. The dependent variables are the next 5-, 10-, and 20-day cumulative returns ($Ret_{k,t+1,t+5}$, $Ret_{k,t+1,t+10}$, and $Ret_{k,t+1,t+20}$). Realized daily volatility ($RDVol_{k,t}$) and realized daily kurtosis ($RDKurt_{k,t}$) are computed using 5-minute returns with equations (26) and (27). A contemporaneous return $Ret_{k,t}$ is included to control for the well-known factors of return reversals and continuations. In addition to daily higher moments, the market size of a firm ($LogME_{k,t}$), the book-to-market ratio ($LogBM_{k,t}$), the cumulative returns over the 11 months from the previous 12-month to the previous 2-month period ($Momentum_{k,t}$), and the percentage of daily trading volume ($Volume(\%)_{k,t}$) are included as control variables. The coefficient estimates and corresponding t-statistics are computed using Fama-MacBeth (1973) regressions with Newey-West standard errors. In the Newey-West standard error correction, a lag of 4, 9, or 19 is applied to the next 5-, 10-, and 20-day returns, respectively. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Information sample			*No-information sample*		
	$Ret_{t+1,t+5}$	$Ret_{t+1,t+10}$	$Ret_{t+1,t+20}$	$Ret_{t+1,t+5}$	$Ret_{t+1,t+10}$	$Ret_{t+1,t+20}$
OrthRDSkew	1.87**	2.03*	1.37	0.69	0.50	-0.66
	(2.55)	(1.88)	(0.92)	(1.41)	(0.75)	(-0.61)
RDVol	456.82**	850.59**	1460.59**	685.36***	1160.22***	2282.04***
	(2.16)	(2.34)	(2.28)	(3.31)	(3.32)	(3.47)
RDKurt	-0.39*	-0.65*	-1.14*	-0.42**	-0.46	-0.90*
	(-1.68)	(-1.76)	(-1.85)	(-2.23)	(-1.46)	(-1.68)
Ret	-0.00	-0.00	0.01	-0.05***	-0.06***	-0.07***
	(-0.77)	(-0.27)	(0.63)	(-12.85)	(-9.66)	(-7.57)
LogME	-6.14***	-10.45***	-19.59***	-6.11***	-11.91***	-21.66***
	(-6.10)	(-5.60)	(-5.43)	(-6.46)	(-6.69)	(-6.17)
LogBM	4.72***	7.86**	14.95**	3.70***	7.08***	14.21***
	(3.01)	(2.53)	(2.37)	(2.99)	(2.85)	(2.78)
Momentum	3.68	-1.55	-10.25	-3.65	-8.43	-15.67
	(0.47)	(-0.10)	(-0.32)	(-0.64)	(-0.73)	(-0.64)
Volume (%)	-217.38**	-420.27***	-901.83***	27.42	-178.87	-602.17**
	(-2.47)	(-3.05)	(-3.70)	(0.38)	(-1.34)	(-2.52)
Constant	67.73***	116.32***	226.14***	59.51***	115.72***	219.78***
	(5.52)	(4.80)	(4.62)	(5.69)	(5.67)	(5.38)
Observations	1622226	1622226	1622226	4269197	4269197	4269197
R^2	0.09	0.09	0.09	0.07	0.06	0.06

Table 38: Realized skewness and future returns: Drift-adjusted realized daily skewness

This table examines the relationship between realized skewness and future stock returns depending on intangible information releases using a modified measure of realized daily skewness to mitigate the effect of daily realized returns. Drift-adjusted realized daily skewness $AdjRDSkew_{k,t}$, volatility $AdjRDVol_{k,t}$, and kurtosis $AdjRDKurt_{k,t}$ are computed using equations (29), (30), and (32), respectively. The dependent variables are the next 5-, 10-, and 20-day cumulative returns ($Ret_{k,t+1,t+5}$, $Ret_{k,t+1,t+10}$, and $Ret_{k,t+1,t+20}$). A contemporaneous return $Ret_{k,t}$ is included to control for the well-known factors of return reversals and continuations. As other control variables, we include the market size of a firm ($LogME_{k,t}$), book-to-market ratio ($LogBM_{k,t}$), cumulative returns over the 11 months from the previous 12-month to the previous 2-month period ($Momentum_{k,t}$), and percentage of daily trading volume ($Volume(\%)_{k,t}$). The coefficient estimates and corresponding t-statistics are computed using Fama-MacBeth (1973) regressions with Newey-West standard errors. In the Newey-West standard error correction, a lag of 4, 9, or 19 is applied to the next 5-, 10-, and 20-day returns, respectively. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Information sample			*No-information sample*		
	$Ret_{t+1,t+5}$	$Ret_{t+1,t+10}$	$Ret_{t+1,t+20}$	$Ret_{t+1,t+5}$	$Ret_{t+1,t+10}$	$Ret_{t+1,t+20}$
AdjRDSkew	2.81***	3.69***	3.53***	0.65*	0.62	0.25
	(4.67)	(4.50)	(3.48)	(1.76)	(1.25)	(0.32)
AdjRDVol	385.30*	750.93**	1332.26**	679.88***	1161.89***	2166.84***
	(1.81)	(2.09)	(2.06)	(3.32)	(3.34)	(3.33)
AdjRDKurt	-0.37	-0.54	-1.09*	-0.42**	-0.45	-0.82
	(-1.60)	(-1.48)	(-1.75)	(-2.20)	(-1.45)	(-1.52)
Ret	0.00	0.00	0.01	-0.05***	-0.05***	-0.06***
	(0.48)	(0.59)	(1.37)	(-12.27)	(-8.80)	(-6.88)
LogME	-6.17***	-10.89***	-20.18***	-6.18***	-11.73***	-21.47***
	(-6.10)	(-6.03)	(-5.84)	(-6.73)	(-6.73)	(-6.21)
LogBM	5.01***	7.69***	13.97**	3.47***	6.76***	13.68***
	(3.32)	(2.72)	(2.38)	(2.95)	(2.86)	(2.80)
Momentum	-0.05	-4.73	-13.12	-3.20	-8.61	-16.67
	(-0.01)	(-0.31)	(-0.42)	(-0.55)	(-0.73)	(-0.66)
Volume (%)	-238.16***	-457.48***	-906.11***	30.70	-160.33	-576.38**
	(-2.73)	(-3.36)	(-3.89)	(0.43)	(-1.21)	(-2.45)
Constant	69.15***	122.25***	236.61***	59.81***	114.33***	219.85***
	(5.68)	(5.17)	(4.95)	(5.85)	(5.65)	(5.39)
Observations	1280100	1280100	1280100	4774766	4774766	4774766
R^2	0.10	0.10	0.10	0.07	0.06	0.06

Table 39: Realized skewness and future returns: Excluding major price shocks

This table examines the relationship between realized skewness and future stock returns depending on intangible information releases after excluding large price changes to mitigate concerns of return reversal or continuation. In Panel A, the firm-day observations with daily returns greater than 10% or less than -10% are excluded. For further robustness, we exclude the observations with daily returns greater than 5% or less than -5% in Panel B. Realized daily skewness ($RDSkew_{k,t}$) is computed from equation (25), and mean-adjusted realized daily skewness ($AdjRDSkew_{k,t}$) is computed using equation (29). The dependent variables are the next 5-, 10-, and 20-day cumulative returns ($Ret_{k,t+1,t+5}$, $Ret_{k,t+1,t+10}$, and $Ret_{k,t+1,t+20}$). The information dummy variable, $d\_Info_{k,t}$, is equal to 1 if a firm-day observation is accompanied by information releases (i.e., analyst earnings forecast or recommendation) and 0 otherwise. A contemporaneous return $Ret_{k,t}$ is included to control for the well-known factors of return reversals and continuations. As other control variables, we include the market size of a firm ($LogME_{k,t}$), book-to-market ratio ($LogBM_{k,t}$), cumulative returns over the 11 months from the previous 12-month to the previous 2-month period ($Momentum_{k,t}$), and percentage of daily trading volume ($Volume(\%)_{k,t}$). The coefficient estimates and corresponding t-statistics are computed using Fama-MacBeth (1973) regressions with Newey-West standard errors. In the Newey-West standard error correction, a lag of 4, 9, or 19 is applied to the next 5-, 10-, and 20-day returns, respectively. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Excluding ±10% changes

	$Ret_{t+1,t+5}$		$Ret_{t+1,t+10}$		$Ret_{t+1,t+20}$	
	(1)	(2)	(3)	(4)	(5)	(6)
RDSkew	1.57***		1.82***		1.66**	
	(4.16)		(3.57)		(2.16)	
RDSkew×d_Info	0.58		1.72**		2.21**	
	(1.06)		(2.28)		(2.14)	
AdjRDSkew		1.18***		1.51***		1.45*
		(3.18)		(3.00)		(1.92)
AdjRDSkew×d_Info		1.49***		2.75***		3.24***
		(2.60)		(3.48)		(3.04)
Ret	-0.06***	-0.05***	-0.07***	-0.06***	-0.08***	-0.08***
	(-12.34)	(-12.16)	(-10.32)	(-10.10)	(-8.63)	(-8.55)
Ret×d_Info	0.05***	0.04***	0.05***	0.05***	0.07***	0.07***
	(12.05)	(12.33)	(10.14)	(10.60)	(8.51)	(9.00)
LogME	-7.97***	-7.99***	-14.70***	-14.72***	-27.10***	-27.12***
	(-6.75)	(-6.75)	(-6.43)	(-6.43)	(-5.98)	(-5.97)
LogBM	2.52*	2.51*	4.63*	4.62*	9.71*	9.70*
	(1.93)	(1.92)	(1.74)	(1.73)	(1.79)	(1.78)
Momentum	-3.49	-3.52	-10.67	-10.71	-21.40	-21.45
	(-0.51)	(-0.51)	(-0.78)	(-0.78)	(-0.74)	(-0.74)
Volume (%)	53.45	54.74	-55.50	-53.19	-370.77	-368.07
	(0.58)	(0.60)	(-0.33)	(-0.31)	(-1.19)	(-1.18)
Constant	87.66***	87.69***	161.56***	161.58***	305.90***	305.86***
	(5.94)	(5.94)	(5.71)	(5.70)	(5.41)	(5.41)
Observations	5975812	5975812	5975812	5975812	5975812	5975812
R^2	0.05	0.05	0.05	0.05	0.05	0.05

Panel A. Excluding ±5% changes						
	$Ret_{t+1,t+5}$		$Ret_{t+1,t+10}$		$Ret_{t+1,t+20}$	
	(1)	(2)	(3)	(4)	(5)	(6)
RDSkew	1.76***		1.88***		1.61**	
	(4.52)		(3.44)		(2.06)	
RDSkew×d_Info	0.87		2.54***		3.00***	
	(1.51)		(2.95)		(2.65)	
AdjRDSkew		1.41***		1.57***		1.40*
		(3.58)		(2.86)		(1.81)
AdjRDSkew×d_Info		1.58***		3.42***		3.87***
		(2.61)		(3.77)		(3.26)
Ret	-0.06***	-0.06***	-0.07***	-0.06***	-0.08***	-0.08***
	(-11.55)	(-11.30)	(-9.13)	(-8.91)	(-7.62)	(-7.54)
Ret×d_Info	0.03***	0.03***	0.04***	0.04***	0.05***	0.05***
	(7.70)	(7.98)	(6.28)	(6.84)	(5.30)	(5.80)
LogME	-7.40***	-7.41***	-13.62***	-13.63***	-25.29***	-25.31***
	(-6.33)	(-6.33)	(-6.08)	(-6.09)	(-5.68)	(-5.68)
LogBM	2.20*	2.19*	4.17	4.16	8.47	8.46
	(1.75)	(1.75)	(1.64)	(1.64)	(1.62)	(1.62)
Momentum	-3.48	-3.52	-10.07	-10.12	-19.20	-19.24
	(-0.50)	(-0.51)	(-0.73)	(-0.73)	(-0.67)	(-0.67)
Volume (%)	37.22	37.89	-72.61	-71.33	-417.17	-416.06
	(0.35)	(0.36)	(-0.37)	(-0.36)	(-1.17)	(-1.17)
Constant	83.54***	83.56***	153.26***	153.28***	291.85***	291.85***
	(5.75)	(5.75)	(5.51)	(5.51)	(5.23)	(5.23)
Observations	5620871	5620871	5620871	5620871	5620871	5620871
R^2	0.05	0.05	0.05	0.05	0.05	0.05

Table 40: Robustness: Liquidity provisions

This table shows the results of robustness tests on realized skewness to the liquidity provision of market makers around news releases. Realized daily skewness ($RDSkew_{k,t}$), realized daily volatility ($RDVol_{k,t}$), and realized daily kurtosis ($RDKurt_{k,t}$) are computed using 5-minute returns with equations (25), (26), and (27), respectively. The dependent variables are the next 5-, 10-, and 20-day cumulative returns ($Ret_{k,t+1,t+5}$, $Ret_{k,t+1,t+10}$, and $Ret_{k,t+1,t+20}$). A contemporaneous return $Ret_{k,t}$ is included to control for the well-known factors of return reversals and continuations. The other main control variable in this table is $Ret_{k,t-3,t-1}$ to control for the liquidity provision of market makers. As other control variables, we include the market size of a firm ($LogME_{k,t}$), book-to-market ratio ($LogBM_{k,t}$), cumulative returns over the 11 months from the previous 12-month to the previous 2-month period ($Momentum_{k,t}$), and percentage of daily trading volume ($Volume(\%)_{k,t}$). The coefficient estimates and corresponding t-statistics are computed using Fama-MacBeth (1973) regressions. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Information sample			*No-information sample*		
	$Ret_{t+1,t+5}$	$Ret_{t+1,t+10}$	$Ret_{t+1,t+20}$	$Ret_{t+1,t+5}$	$Ret_{t+1,t+10}$	$Ret_{t+1,t+20}$
RDSkew	1.90***	2.52***	2.16**	0.81**	0.63	0.10
	(3.38)	(3.48)	(2.20)	(2.40)	(1.40)	(0.16)
RDVol	384.97***	742.85***	1287.29***	690.64***	1162.55***	2126.93***
	(2.96)	(4.46)	(5.74)	(6.07)	(8.16)	(11.10)
RDKurt	-0.40**	-0.57***	-1.06***	-0.41***	-0.47***	-0.79***
	(-2.56)	(-2.91)	(-4.03)	(-3.75)	(-3.32)	(-4.32)
Ret	0.00	0.00	0.01	-0.05***	-0.06***	-0.06***
	(0.29)	(0.28)	(1.52)	(-14.45)	(-11.68)	(-10.33)
$Ret_{t-3,t-1}$	-0.00	-0.00	0.00	-0.02***	-0.02***	-0.03***
	(-1.25)	(-0.83)	(0.62)	(-9.65)	(-8.19)	(-7.52)
LogME	-6.19***	-10.85***	-19.98***	-6.24***	-11.85***	-21.57***
	(-9.90)	(-13.14)	(-17.32)	(-11.83)	(-16.71)	(-21.19)
LogBM	4.66***	7.24***	13.61***	3.06***	6.34***	13.03***
	(5.23)	(5.83)	(7.60)	(4.83)	(7.08)	(9.88)
Momentum	-0.92	-6.42	-13.73	-4.01	-9.81**	-16.77***
	(-0.21)	(-1.06)	(-1.58)	(-1.32)	(-2.23)	(-2.60)
Volume (%)	-241.77***	-468.60***	-912.17***	40.35	-152.15**	-568.17***
	(-3.96)	(-5.89)	(-8.85)	(0.95)	(-2.56)	(-7.24)
Constant	69.81***	123.27***	235.18***	60.29***	116.25***	220.24***
	(9.70)	(12.36)	(16.22)	(10.53)	(14.60)	(18.99)
Observations	1279953	1279953	1279953	4773072	4773072	4773072
R^2	0.11	0.11	0.10	0.07	0.07	0.07

APPENDIX A

MISCELLANEOUS SECTION FOR CHAPTER 1

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 [73]. I follow the procedure and notations prescribed in his paper. [73] 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:

$$p_t = m_t + s_t. \tag{33}$$

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 [73] in which he estimates the following vector autoregression system with five lags:

$$\begin{aligned} r_t &= a_1 r_{t-1} + a_2 r_{t-2} + \cdots + b_1 x_{t-1} + b_2 x_{t-2} + \cdots + v_{1,t}, \\ x_t &= c_1 r_{t-1} + c_2 r_{t-2} + \cdots + d_1 x_{t-1} + d_2 x_{t-2} + \cdots + v_{2,t}, \end{aligned} \tag{34}$$

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 [73], 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:

$$\begin{aligned} r_t &= a_0^* v_{1,t} + a_1^* v_{1,t-1} + a_2^* v_{1,t-2} + \cdots + b_0^* v_{2,t} + b_1^* v_{2,t-1} + b_2^* v_{2,t-2} + \cdots \\ x_t &= c_0^* v_{1,t} + c_1^* v_{1,t-1} + c_2^* v_{1,t-2} + \cdots + d_0^* v_{2,t} + d_1^* v_{2,t-1} + d_2^* v_{2,t-2} + \cdots \end{aligned} \tag{35}$$

Using Equation (35) and the identification restriction of [19], the pricing error can be expressed as

$$s_t = \alpha_0 v_{1,t} + \alpha_1 v_{1,t-1} + \cdots + \beta_0 v_{2,t} + \beta_1 v_{2,t-1} + \cdots , \tag{36}$$

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

$$\sigma^2(s) = \sum_{j=0}^{\infty} [\alpha_j, \beta_j] Cov(v) \begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix} . \tag{37}$$

In the implementation, I use transaction data in the TAQ database and a filter used in [20]. To assign trade direction, I use the algorithm of [89]. 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. [67] examine various measures of liquidity and their performance in capturing the price impact, and find that a measure introduced by [5] 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 [5] and zeros defined as the proportion of positive days with zero-returns.

The measure of illiquidity introduced by [5] 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:

$$Amihud_{i,t} = \frac{1}{\#of\ Days} \sum_{d=1}^{\#of\ Days} \frac{|R^i_{t,d}|}{V^i_{t,d}}, \tag{38}$$

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.

[67] 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

$$Zeros_{i,t} = \frac{(\#\text{ of positive-volume days with zero returns of stock } i \text{ in month } t\ )}{(\#\text{ of trading days in month } t)}. \tag{39}$$

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

APPENDIX B

MISCELLANEOUS SECTION FOR CHAPTER 2

B.1 A List of Variables and Their Construction

In this appendix, we provide additional information about variables used in the paper.

- *ME*: Market capitalization of firms as the product of the closing prices and the number of shares outstanding of securities obtained from CRSP.

- *BM*: Ratio of book equity (as of December of the previous calendar year obtained from COMPUSTAT) to market capitalization (CRSP).

- *vol*: Trading volumes of securities obtained from CRSP.

- *Amihud*: Illiquidity measure of [5] constructed as

$$Amihud_{i,t} = \sum_{k=1}^{M} \frac{|r_{i,t,k}|}{dvol_{i,t,k}},$$

  where $|r_{i,t,k}|$ is the k-th 5-minute return of firm i on day t, $dvol_{i,t,k}$ is a dollar trading volume ($price_{i,t,k} \times volume_{i,t,k}$) over the k-th 5-minute interval on day t for firm i, and $M = 78$ because of 5-minute sampling frequency in a day using TAQ database.

- *Stock Spread(%)*: Percentage spread of bid and ask prices computed as

$$Stock\ Spread(\%)_{i,t} = 100 \times \frac{Ask_{i,t} - Bid_{i,t}}{0.5(Ask_{i,t} + Bid_{i,t})},$$

  where $Bid_{i,t}$ and $Ask_{i,t}$ are bid and ask prices of security i on day t (CRSP).

- *O/S ratio*: Ratio of aggregate trading volume of listed options (obtained from OptionMetrics) of an underlying security to trading volume of the security (obtained from CRSP).

- *Slope_IVS*: Slope of implied volatilities of traded options estimated from the regression of implied volatilities on their deltas (with deltas of put options being reversed in sign). Implied volatilities and deltas of options are obtained from OptionMetrics.

- *Option Spread(%)*: Percentage spread as the average bid-ask spread divided by the midpoint of bid and ask prices over all options traded. Bid and ask prices of options are obtained from OptionMetrics.

- *Delta*: Average of deltas of options obtained from OptionMetrics (with deltas of put options being reversed in sign).

- *Analyst Disp*: Standard deviation of analyst earnings forecasts obtained from the Institutional Brokers Estimate System (I/B/E/S).

- *IO ratio*: Fraction of a firm's outstanding shares held by institutions obtained from CRSP and Thomson Financial.

- *avg Ivol*: Average of implied volatilities of options traded. Implied volatilities of options are obtained from OptionMetrics.

REFERENCES

[1] Aït-Sahalia, Y., Mykland, P. A., and Zhang, L., "How often to sample a continuous-time process in the presence of market microstructure noise," *Review of Financial Studies*, vol. 18, no. 2, pp. 351–416, 2005.

[2] Albuquerque, R., "Skewness in stock returns: Reconciling the evidence on firm versus aggregate returns," *Review of Financial Studies*, vol. 25, no. 5, pp. 1630–1673, 2012.

[3] Altinkiliç, O. and Hansen, R. S., "On the information role of stock recommendation revisions," *Journal of Accounting and Economics*, vol. 48, no. 1, pp. 17–36, 2009.

[4] Amaya, D., Christoffersen, P., Jacobs, K., and Vasquez, A., "Does realized skewness predict the cross-section of equity returns?," *Journal of Financial Economics*, vol. 118, no. 1, pp. 135–167, 2015.

[5] Amihud, Y., "Illiquidity and stock returns: cross-section and time-series effects," *Journal of Financial Markets*, vol. 5, pp. 31–56, January 2002.

[6] Amihud, Y. and Mendelson, H., "Asset pricing and the bid-ask spread," *Journal of Financial Economics*, vol. 17, no. 2, pp. 223–249, 1986.

[7] Andersen, T. G., "Return volatility and trading volume: An information flow interpretation of stochastic volatility," *The Journal of Finance*, vol. 51, no. 1, pp. 169–204, 1996.

[8] Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H., "The distribution of realized stock return volatility," *Journal of Financial Economics*, vol. 61, no. 1, pp. 43–76, 2001.

[9] Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P., "Modeling and forecasting realized volatility," *Econometrica*, vol. 71, no. 2, pp. 579–625, 2003.

[10] Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X., "The cross-section of volatility and expected returns," *The Journal of Finance*, vol. 61, no. 1, pp. 259–299, 2006.

[11] Appel, I. R., Gormley, T. A., and Keim, D. B., "Passive investors, not passive owners," *Journal of Financial Economics*, vol. 121, no. 1, pp. 111–141, 2016.

[12] Ball, R. and Brown, P., "An empirical evaluation of accounting income numbers," *Journal of Accounting Research*, vol. 6, no. 2, pp. 159–178, 1968.

[13] Bandi, F. M. and Russell, J. R., "Separating microstructure noise from volatility," *Journal of Financial Economics*, vol. 79, no. 3, pp. 655–692, 2006.

[14] Bandi, F. M. and Russell, J. R., "Microstructure noise, realized variance, and optimal sampling," *Review of Economic Studies*, vol. 75, no. 2, pp. 339–369, 2008.

[15] Barndorff-Nielsen, O. E. and Shephard, N., "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, pp. 253–280, 2002.

[16] Beaver, W. H., "The information content of annual earnings announcements," *Journal of Accounting Research*, pp. 67–92, 1968.

[17] Ben-David, I., Franzoni, F., and Moussawi, R., "Do ETFs Increase Volatility?," NBER Working Papers 20071, National Bureau of Economic Research, Inc, Apr. 2014.

[18] Bernard, V. L. and Thomas, J. K., "Post-earnings-announcement drift: Delayed price response or risk premium?," *Journal of Accounting Research*, vol. 27, pp. 1–36, 1989.

[19] Beveridge, S. and Nelson, C. R., "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*, vol. 7, no. 2, pp. 151–174, 1981.

[20] Boehmer, E. and Wu, J. J., "Short Selling and the Price Discovery Process," *Review of Financial Studies*, vol. 26, no. 2, pp. 287–322, 2013.

[21] Boni, L. and Womack, K. L., "Analysts, industries, and price momentum," *Journal of Financial and Quantitative Analysis*, vol. 41, no. 01, pp. 85–109, 2006.

[22] Boone, A. L. and White, J. T., "The effect of institutional ownership on firm transparency and information production," *Journal of Financial Economics*, vol. 117, no. 3, pp. 508–533, 2015.

[23] Bradley, D., Clarke, J. E., Lee, S. S., and Ornthanalai, C., "Are analysts recommendations informative? intraday evidence on the impact of time stamp delays," *The Journal of Finance*, vol. 69, no. 2, pp. 645–673, 2013.

[24] Brav, A., Jiang, W., Partnoy, F., and Thomas, R., "Hedge Fund Activism, Corporate Governance, and Firm Performance," *Journal of Finance*, vol. 63, pp. 1729–1775, 08 2008.

[25] Bremer, M. and Sweeney, R. J., “The reversal of large stock-price decreases,” *The Journal of Finance*, vol. 46, no. 2, pp. 747–754, 1991.

[26] Brennan, M. J. and Subrahmanyam, A., “Market microstructure and asset pricing: On the compensation for illiquidity in stock returns,” *Journal of Financial Economics*, vol. 41, no. 3, pp. 441–464, 1996.

[27] Busse, J. A. and Green, C. T., “Market efficiency in real time,” *Journal of Financial Economics*, vol. 65, pp. 415–437, September 2002.

[28] Calonico, S., Cattaneo, M. D., and Titiunik, R., “Robust Nonparametric Confidence Intervals for RegressionDiscontinuity Designs,” *Econometrica*, vol. 82, pp. 2295–2326, November 2014.

[29] Cao, C., Liang, B., Lo, A. W., and Petrasek, L., “Hedge fund holdings and stock market efficiency,” *The Review of Asset Pricing Studies*, 2014.

[30] Carhart, M. M., “On persistence in mutual fund performance,” *The Journal of Finance*, vol. 52, no. 1, pp. 57–82, 1997.

[31] Chae, J., “Trading volume, information asymmetry, and timing information,” *The Journal of Finance*, vol. 60, no. 1, pp. 413–442, 2005.

[32] Chakravarty, S., Gulen, H., and Mayhew, S., “Informed trading in stock and option markets,” *Journal of Finance*, vol. 59, no. 3, pp. 1235–1257, 2004.

[33] Chan, W. S., “Stock price reaction to news and no-news: Drift and reversal after headlines,” *Journal of Financial Economics*, vol. 70, no. 2, pp. 223–260, 2003.

[34] Chang, Y.-C., Hong, H., and Liskovich, I., “Regression Discontinuity and the Price Effects of Stock Market Indexing,” *Review of Financial Studies*, vol. 28, no. 1, pp. 212–246, 2015.

[35] Chen, E.-T. J. and Clements, A., “S&P 500 implied volatility and monetary policy announcements,” *Finance Research Letters*, vol. 4, pp. 227–232, December 2007.

[36] Chen, J., Hong, H., and Stein, J. C., “Forecasting crashes: Trading volume, past returns, and conditional skewness in stock prices,” *Journal of Financial Economics*, vol. 61, no. 3, pp. 345–381, 2001.

[37] Chen, Q., Goldstein, I., and Jiang, W., “Price informativeness and investment sensitivity to stock price,” *Review of Financial Studies*, vol. 20, no. 3, pp. 619–650, 2007.

[38] Chordia, T., Roll, R., and Subrahmanyam, A., “Evidence on the speed of convergence to market efficiency,” *Journal of Financial Economics*, vol. 76, pp. 271–292, May 2005.

[39] Conrad, J. S., Hameed, A., and Niden, C., "Volume and autocovariances in short-horizon individual security returns," *The Journal of Finance*, vol. 49, no. 4, pp. 1305–1329, 1994.

[40] Copeland, T. E. and Galai, D., "Information effects on the bid-ask spread," *The Journal of Finance*, vol. 38, no. 5, pp. 1457–1469, 1983.

[41] Cox, D. R. and Peterson, D. R., "Stock returns following large one-day declines: Evidence on short-term reversals and longer-term performance," *Journal of Finance*, vol. 49, no. 1, pp. 255–267, 1994.

[42] Crane, A. D., Michenaud, S., and Weston, J. P., "The effect of institutional ownership on payout policy: Evidence from index thresholds," *The Review of Financial Studies*, vol. 29, no. 6, p. 1377, 2016.

[43] Da, Z. and Shive, S., "Exchange traded funds and asset return correlations," *NBER Working Paper*, 2016.

[44] Donders, M., Kouwenberg, R., and Vorst, T., "Options and earnings announcements: an empirical study of volatility, trading volume, open interest and liquidity," *European Financial Management*, vol. 6, no. 2, pp. 149–171, 2000.

[45] Donders, M. W. and Vorst, T. C., "The impact of firm specific news on implied volatilities," *Journal of Banking & Finance*, vol. 20, no. 9, pp. 1447–1461, 1996.

[46] Easley, D., Kiefer, N. M., O'hara, M., and Paperman, J. B., "Liquidity, information, and infrequently traded stocks," *The Journal of Finance*, vol. 51, no. 4, pp. 1405–1436, 1996.

[47] Easley, D., O'hara, M., and Srinivas, P. S., "Option volume and stock prices: Evidence on where informed traders trade," *Journal of Finance*, vol. 53, no. 2, pp. 431–465, 1998.

[48] Ederington, L. H. and Lee, J. H., "The creation and resolution of market uncertainty: the impact of information releases on implied volatility," *Journal of Financial and Quantitative Analysis*, vol. 31, no. 04, pp. 513–539, 1996.

[49] Elton, E. J., Gruber, M. J., and Blake, C. R., "Holdings Data, Security Returns, and the Selection of Superior Mutual Funds," *Journal of Financial and Quantitative Analysis*, vol. 46, pp. 341–367, April 2011.

[50] Elton, E. J., Gruber, M. J., Das, S., and Hlavka, M., "Efficiency with costly information: A reinterpretation of evidence from managed portfolios," *Review of Financial Studies*, vol. 6, no. 1, pp. 1–22, 1993.

[51] Epstein, L. G. and Schneider, M., "Ambiguity, information quality, and asset pricing," *The Journal of Finance*, vol. 63, no. 1, pp. 197–228, 2008.

[52] Falkenstein, E. G., " Preferences for Stock Characteristics as Revealed by Mutual Fund Portfolio Holdings," *Journal of Finance*, vol. 51, pp. 111–135, March 1996.

[53] Fama, E. F., "Random walks in stock market prices," *Financial Analysts Journal*, vol. 51, no. 1, pp. 75–80, 1995.

[54] Fama, E. F. and French, K. R., "The cross-section of expected stock returns," *The Journal of Finance*, vol. 47, no. 2, pp. 427–465, 1992.

[55] Fama, E. F. and French, K. R., "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, vol. 33, no. 1, pp. 3–56, 1993.

[56] Fama, E. F. and French, K. R., "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, vol. 33, no. 1, pp. 3–56, 1993.

[57] Fama, E. F. and French, K. R., "Luck versus Skill in the Cross-Section of Mutual Fund Returns," *Journal of Finance*, vol. 65, pp. 1915–1947, October 2010.

[58] Fama, E. F. and MacBeth, J. D., "Risk, return, and equilibrium: Empirical tests," *The Journal of Political Economy*, vol. 81, no. 3, pp. 607–636, 1973.

[59] Francis, J., Schipper, K., and Vincent, L., "Earnings announcements and competing information," *Journal of Accounting and Economics*, vol. 33, no. 3, pp. 313–342, 2002.

[60] Frankel, R., Kothari, S., and Weber, J., "Determinants of the informativeness of analyst research," *Journal of Accounting and Economics*, vol. 41, no. 1, pp. 29–54, 2006.

[61] Frazzini, A. and Lamont, O. A., "Dumb money: Mutual fund flows and the cross-section of stock returns," *Journal of Financial Economics*, vol. 88, pp. 299–322, May 2008.

[62] Frino, A. and Gallagher, D. R., "Tracking s&p 500 index funds," *Journal of Portfolio Management*, vol. 28, no. 1, pp. 44–55, 2001.

[63] Glosten, L. R. and Harris, L. E., "Estimating the components of the bid/ask spread," *Journal of Financial Economics*, vol. 21, no. 1, pp. 123–142, 1988.

[64] Glosten, L. R. and Milgrom, P. R., "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders," *Journal of Financial Economics*, vol. 14, no. 1, pp. 71–100, 1985.

[65] Glosten, L. R., Nallareddy, S., and Zou, Y., "Etf trading and informational efficiency of underlying securities," *SSRN Working Paper*, 2016.

[66] Gormley, T. A. and Matsa, D. A., "Common errors: How to (and not to) control for unobserved heterogeneity," *Review of Financial Studies*, vol. 27, no. 2, pp. 617–661, 2014.

[67] Goyenko, R. Y., Holden, C. W., and Trzcinka, C. A., "Do liquidity measures measure liquidity?," *Journal of Financial Economics*, vol. 92, pp. 153–181, May 2009.

[68] Griffin, J. M., Kelly, P. J., and Nardari, F., "Do Market Efficiency Measures Yield Correct Inferences? A Comparison of Developed and Emerging Markets," *Review of Financial Studies*, vol. 23, pp. 3225–3277, August 2010.

[69] Grossman, S. J. and Stiglitz, J. E., "On the Impossibility of Informationally Efficient Markets," *American Economic Review*, vol. 70, pp. 393–408, June 1980.

[70] Hahn, J., Todd, P., and Van der Klaauw, W., "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, vol. 69, pp. 201–209, January 2001.

[71] Hamm, S. J., "The effect of etfs on stock liquidity," *SSRN Working Paper*, 2014.

[72] Harvey, C. R. and Siddique, A., "Conditional skewness in asset pricing tests," *The Journal of Finance*, vol. 55, no. 3, pp. 1263–1295, 2000.

[73] Hasbrouck, J., "Assessing the Quality of a Security Market: A New Approach to Transaction-Cost Measurement," *Review of Financial Studies*, vol. 6, no. 1, pp. 191–212, 1993.

[74] Hou, K. and Moskowitz, T. J., "Market Frictions, Price Delay, and the Cross-Section of Expected Returns," *Review of Financial Studies*, vol. 18, no. 3, pp. 981–1020, 2005.

[75] Hu, J., "Does option trading convey stock price information?," *Journal of Financial Economics*, vol. 111, no. 3, pp. 625–645, 2014.

[76] Iliev, P. and Lowry, M., "Are Mutual Funds Active Voters?," *Review of Financial Studies*, vol. 28, no. 2, pp. 446–485, 2015.

[77] Illeditsch, P. K., "Ambiguous Information, Portfolio Inertia, and Excess Volatility," *Journal of Finance*, vol. 66, pp. 2213–2247, December 2011.

[78] Ippolito, R. A., "Efficiency with Costly Information: A Study of Mutual Fund Performance, 19651984," *The Quarterly Journal of Economics*, vol. 104, no. 1, pp. 1–23, 1989.

[79] Isakov, D. and Perignon, C., "Evolution of market uncertainty around earnings announcements," *Journal of Banking & Finance*, vol. 25, pp. 1769–1788, September 2001.

[80] Israeli, D., Lee, C. M. C., and Sridharan, S. A., "Is there a dark side to exchange traded funds? an information perspective," *Review of Accounting Studies*, vol. 22, pp. 1048–1083, Sep 2017.

[81] Ivković, Z. and Jegadeesh, N., "The timing and value of forecast and recommendation revisions," *Journal of Financial Economics*, vol. 73, no. 3, pp. 433–463, 2004.

[82] Jegadeesh, N. and Kim, W., "Do analysts herd? an analysis of recommendations and market reactions," *Review of Financial Studies*, vol. 23, no. 2, pp. 901–937, 2010.

[83] Jegadeesh, N. and Titman, S., "Returns to buying winners and selling losers: Implications for stock market efficiency," *The Journal of Finance*, vol. 48, no. 1, pp. 65–91, 1993.

[84] Jiang, G. J., Xu, D., and Yao, T., "The information content of idiosyncratic volatility," *Journal of Financial and Quantitative Analysis*, vol. 44, no. 01, pp. 1–28, 2009.

[85] Johnson, T. L. and So, E. C., "The option to stock volume ratio and future returns," *Journal of Financial Economics*, vol. 106, no. 2, pp. 262–286, 2012.

[86] Kelly, B., Pástor, L., and Veronesi, P., "The Price of Political Uncertainty: Theory and Evidence from the Option Market," *Journal of Finance*, vol. 71, pp. 2417–2480, October 2016.

[87] Kokkonen, J. and Suominen, M., "Hedge Funds and Stock Market Efficiency," *Management Science*, vol. 61, pp. 2890–2904, December 2015.

[88] Lee, C. and Swaminathan, B., "Price momentum and trading volume," *The Journal of Finance*, vol. 55, no. 5, pp. 2017–2069, 2000.

[89] Lee, C. M. C. and Ready, M. J., " Inferring Trade Direction from Intraday Data," *Journal of Finance*, vol. 46, pp. 733–746, June 1991.

[90] Lee, D. S. and Lemieux, T., "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, vol. 48, pp. 281–355, June 2010.

[91] Leippold, M., Trojani, F., and Vanini, P., "Learning and Asset Prices Under Ambiguous Information," *Review of Financial Studies*, vol. 21, pp. 2565–2597, November 2008.

[92] Loh, R. K. and Stulz, R. M., "When are analyst recommendation changes influential?," *The Review of Financial Studies*, vol. 24, no. 2, pp. 593–627, 2011.

[93] Moeller, S. B., Schlingemann, F. P., and Stulz, R. M., "How do diversity of opinion and information asymmetry affect acquirer returns?," *Review of Financial Studies*, vol. 20, no. 6, pp. 2047–2078, 2007.

[94] Nagel, S., "Evaporating liquidity," *Review of Financial Studies*, vol. 25, no. 7, pp. 2005–2039, 2012.

[95] Neuberger, A., "Realized skewness," *Review of Financial Studies*, vol. 25, no. 11, pp. 3423–3455, 2012.

[96] Pástor, L. and Stambaugh, R. F., "On the Size of the Active Management Industry," *Journal of Political Economy*, vol. 120, no. 4, pp. 740–781, 2012.

[97] Pritamani, M. and Singal, V., "Return predictability following large price changes and information releases," *Journal of Banking and Finance*, vol. 25, no. 4, pp. 631–656, 2001.

[98] Roll, R., Schwartz, E., and Subrahmanyam, A., "O/s: The relative trading activity in options and stock," *Journal of Financial Economics*, vol. 96, no. 1, pp. 1–17, 2010.

[99] Rösch, D. M., Subrahmanyam, A., and Van Dijk, M. A., "The dynamics of market efficiency," *Review of Financial Studies*, vol. 30, no. 4, pp. 1151–1187, 2016.

[100] Sadka, R., "Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk," *Journal of Financial Economics*, vol. 80, no. 2, pp. 309–349, 2006.

[101] Saffi, P. A. C. and Sigurdsson, K., "Price Efficiency and Short Selling," *Review of Financial Studies*, vol. 24, no. 3, pp. 821–852, 2011.

[102] Savor, P. G., "Stock returns after major price shocks: The impact of information," *Journal of Financial Economics*, vol. 106, no. 3, pp. 635–659, 2012.

[103] Stock, J. H. and Yogo, M., "Testing for weak instruments in linear iv regression," *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, p. 80, 2005.

[104] Stoll, H. R., "Presidential address: friction," *Journal of Finance*, vol. 55, no. 4, pp. 1479–1514, 2000.

[105] Tetlock, P. C., "Does public financial news resolve asymmetric information?," *Review of Financial Studies*, vol. 23, no. 9, pp. 3520–3557, 2010.

[106] Venkatesh, P. and Chiang, R., "Information asymmetry and the dealer's bid-ask spread: A case study of earnings and dividend announcements," *The Journal of Finance*, vol. 41, no. 5, pp. 1089–1102, 1986.

[107] Womack, K. L., “Do brokerage analysts’ recommendations have investment value?,” *Journal of Finance*, vol. 51, no. 1, pp. 137–167, 1996.

[108] Zhang, L., Mykland, P. A., and Aït-Sahalia, Y., “A tale of two time scales: Determining integrated volatility with noisy high-frequency data,” *Journal of the American Statistical Association*, vol. 100, no. 472, pp. 1394–1411, 2005.

[109] Zhang, X., “Information uncertainty and stock returns,” *The Journal of Finance*, vol. 61, no. 1, pp. 105–137, 2006.

VITA

Youngmin Choi completed his Ph.D. in finance at the Scheller College of Business at Georgia Institute of Technology, and will be joining Baruch College, the Citi University of New York, as an Assistant Professor in Finance in August 2018. Prior to that he received a Master's degree in Quantitative and Computational Finance at Georgia Tech and a Bachelor's degree in Mechanical Engineering and Business Administration from Yonsei University. He also has work experience in the financial industry as an assistant manager in Hedge Fund Investment Group at Woori Investment Securities and Co.

Youngmin's research interests are in empirical asset pricing and derivatives. He has presented his research papers at various finance conferences such as Financial Management Association (FMA), Northern Finance Association (NFA), and Midwest Finance Association (MFA).

Youngmin has also taught undergraduate courses on the investment at Georgia Tech. At Baruch College, he will be teaching an advanced course on corporate finance.