Three Essays on Mispricing and Market Efficiency

Nan Qin

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Business, Finance

Vijay Singal
Gregory B. Kadlec
Raman Kumar
Arthur J. Keown

July 1, 2014
Blacksburg, VA

Keywords: indexing, index funds, ETFs, passive institutional investors, stock price efficiency, passive trading, trading strategy, idiosyncratic risk, equally-weighted index, stock return bias, Jensen's inequality

Three Essays on Mispricing and Market Efficiency

Nan Qin

ABSTRACT

This dissertation consists of three essays. The first essay studies the impact of indexing on stock price efficiency. Indexing has experienced substantial growth over the last two decades because it is an effective way of holding a diversified portfolio while minimizing trading costs and taxes. In this paper, we focus on one negative externality of indexing: the effect on efficiency of stock prices. Based on a sample of large and liquid U.S. stocks, we find that greater indexing leads to less efficient stock prices, as indicated by stronger post-earnings-announcement drift, greater deviations of stock prices from the random walk and greater return predictability from lagged order imbalances. We conjecture that reduced incentives for information acquisition and arbitrage induced by indexing are probably the main cause of the degradation in price efficiency, but we find no evidence supporting a direct impact from passive trading or any effect through liquidity.

The second essay investigates the effect of price inefficiency on idiosyncratic risk and stock returns. I finds that price inefficiency in individual stocks contributes to expected idiosyncratic volatility. If idiosyncratic risk is priced, greater price inefficiency could be associated with higher expected returns. Consistent with this hypothesis, this paper then finds a positive relation between price inefficiency and future stock returns. This return premium of price inefficiency is not explained by traditional risk factors, illiquidity, or transactions costs. It is also evidently different from the return bias related to Jensen's inequality. This paper thus provides new insights about the determinants of expected stock returns, and new supporting evidence that idiosyncratic risk is priced.

The third essay examines whether the upward return bias generated by Jensen's inequality could lead to better performance of equally-weighted (EW) indexes than value-weighted (VW) index when stock prices are not fully efficient. We find that, for a wide range of U.S. stock indexes, EW indexes deliver better four-factor adjusted returns than VW ones do even after deducting transaction costs. Consistent with our hypothesis that the outperformance of EW indexes comes from mispricing, we find that this outperformance concentrates in stocks with greater mispricing, as measured by deviation of stock prices from random walk. Findings in this essay not only imply a potentially winning investment strategy, but also provide new insight into a long-term debate on causes of the outperformance of the EW indexes.

Acknowledgements

I am greatly indebted to the chair of my dissertation committee, Vijay Singal. His meticulous guidance, vigorous support and unwavering commitment not only ensured the successful completion of this dissertation but also greatly helped me in the job market. His attitude towards research and teaching inspire me to become a knowledgeable, professional and successful professor in the future. His fascinating personality has greatly influenced my future career and life.

I would like to thank my other committee members. Gregory B. Kadlec and Raman Kumar shared their valuable experience in academic research and gave me insightful comments and suggestions over the past five years. Raman Kumar and Arthur J. Keown provided me significant support for the job market. I am grateful to Yong Chen for his guidance and help in the first two years of my study in the Ph.D. program and my first research paper. I also appreciate the advice and help that I received from Dilip K. Shome, John C. Easterwood, Douglas M. Patterson, and Ozgur Ince throughout my doctoral study.

I would like to thank my girlfriend, Di Wang, for her company and support during the past one and a half years. I also thank my colleague, Wei-Hsien Li, for being supportive and helpful over the past few years.

My deepest appreciation go to my father, Yu Qin, and my mother, Xiqian Zhi. Their love, understanding and unconditional support were crucial for the completion of my Ph.D. study. They gave me encouragement in overcoming my challenges and obstacles. I dedicate this work to them.

Table of contents

List of Tables

Introduction

Modern asset pricing theories imply no mispricing in equilibrium. However, as suggested by Grossman and Stiglitz (1980), informationally efficient markets are impossible as long as arbitrage is costly. Mispricing, as a result, could be pervasive and profoundly impact capital markets.

Understanding the causes of mispricing is the first and probably the most important step towards a more efficient market. Previous studies have found that price efficiency is affected by liquidity, institutional ownership, investor sentiment, limited arbitrage, underreaction and overreaction, yet little has been explored about whether price efficiency is significantly affected by indexed investments, which have become a significant portion of today's equity market and are experiencing rapid growth. Relative to their active peers, passive institutional investors lack incentives for active information acquisition and price discovery, which are crucial to market efficiency. Increase in passive investors and the consequent reduction in active traders can result in a proportionate increase in information costs borne by each active trader and has the potential to move equilibrium to a less efficient level. The first essay addresses this concern, and finds evidence supporting a negative impact of indexing on stock price efficiency.

Given the difficulty in precisely identifying and quantifying mispricing, Black (1986) suggests that "we are forced to act largely in the dark". Yet it does not imply that investors can do nothing to protect themselves in a market with price inefficiency. Investors could infer the presence and severity of inefficiency from the magnitude of the deviation of stock price from random walk or from the severity of market anomalies such as post earnings announcement drift or stock price comovement. Analyst forecast dispersion, disagreement in investors' opinions, information disclosure and proportion of institutional investors are also likely to be related to price inefficiency. Understanding that mispricing implies greater idiosyncratic uncertainty in future stock returns, underdiversified investors may demand return premiums for holding potentially mispriced stocks. Thus the magnitude of mispricing could be positively associated with expected stock returns. The second essay investigates this hypothesis and finds affirmative evidence. Furthermore, it provides new evidence that idiosyncratic risk is priced, and implies a potentially profitable trading strategy for investors who are able to diversify idiosyncratic risk.

In the first essay, we found that indexing can negatively affect price efficiency. On the other hand, possible mispricing of stocks could result in an upward bias in returns of an equally-weighted index due to Jensen's inequality. This implies that investors could generate alphas by creating EW portfolios based on mispricing of stocks in the index, even if that mispricing is symmetric between large and small cap stocks. Moreover, when the mispricing is asymmetric and greater for high valued stocks, the gains to an equally weighted index is even greater. These hypotheses are tested using simulated stock returns and actual historical returns. The third essay develops and tests these hypotheses using a variety of different indices and different rebalancing periods.

Essays 1 and 3 are co-authored by Nan Qin and Vijay Singal. Essay 2 is sole-authored by Nan Qin.

Essay 1: Indexing and Stock Price Efficiency

1. Introduction

The indexed investment sector, including index mutual funds, enhanced index funds, exchange-traded funds (ETFs), and closet indexers,[1] has experienced rapid growth over the past two decades. As of December 2012, the market share of broadly diversified index funds had reached 27.8% of the total investment in U.S. equities in the mutual fund sector including ETFs.[2] The potential impact of indexing on the efficiency of equity markets, however, is an important but unexplored topic.

Previous studies have studied the impact of indexing on constituent stocks. In their model of institutional investors and index benchmarks, Basak and Pavlova (2013) find that institutional trades amplify stock market volatility and induce excess correlations among index stocks. The implications of the model are supported by Greenwood and Thesmar (2011) who find that institutional ownership increases volatility especially among markets or indexes. Another section of the literature examines the impact of institutional investors on stock prices[3]. Using a comprehensive sample of NYSE-listed stocks between 1983 and 2004, Boehmer and Kelley (2009) find that trading and ownership by institutional investors increase price efficiency. Specifically, institutional investors incorporate information into stock prices through their trading whereas their ownership facilitates informed arbitrage. Based on NYSE/AMEX stocks over 1989-1993, Bartov, Radhakrishnan, and Krinsky (2000) find that institutional holdings are negatively associated with the magnitude of post-earnings-announcement drift, suggesting that investor sophistication may reduce the predictability in stock returns. On the other hand, Chen, Noronha, and Singal (2004) document that stock prices increase immediately after announcements of additions to the S&P 500 index but reverse substantially after 3 months. They explain these findings by

[1] Closet indexers are ‘active’ mutual funds who actually track an index.
[2] Source: 2013 ICI Factbook. The total amount in broad-based equity ETFs ($509 billion) and U.S. equity index mutual funds ($870 billion) was $1,379 billion. The total amount in U.S. equity was $4,965 billion comprising $644 billion in ETFs and $4,320 billion in mutual funds.
[3] Shu (2007) finds that price anomalies, including return momentum, post earnings-announcement drift, and value premium are much stronger in stocks with relatively low institutional trading volume.

the temporary price-pressure caused by index fund purchases of stocks newly added to the index. Goetzmann and Massa (2003) examine daily flows for three major S&P 500 index funds and find a strong contemporaneous correlation between inflows and returns. Similarly, Keim and Madhavan (1997) and Jones and Lipson (1999, 2001) find that index funds generate a larger price impact relative to active funds during the short period following their trading.

This essay is at the intersection of these two literatures: the impact of indexing and the role of institutional investors in enhancing price efficiency. Despite extensive literatures on indexing and price efficiency, there has been no systematic study about the impact of passive index investors on informational efficiency of stock prices. Compared to their active peers, passive institutional investors have two unique characteristics. First, they generally hold a basket of stocks in certain indices passively, without active information acquisition and price discovery. Second, trading of passive funds is mostly driven by investor flows or index changes instead of private information. Both features raise concerns regarding a possible negative impact on price efficiency. As suggested by Grossman and Stiglitz (1980), price discovery relies on informed traders who actively acquire information and incorporate that information into stock prices by trading. An increase in passive (uninformed) investors and the consequent reduction in active traders can result in a proportionate increase in information costs and has the potential to move equilibrium to a less efficient level. In spirit, our essay comes closest to Wurgler (2011) who points to the growing importance of indexing for investment and benchmarking such that it can distort markets and affect the real economy. In contrast, our focus is on a different consequence of the growth in indexing: assimilation of new information into stock prices, which is critical for an efficient market.

We empirically investigate the relation between indexed holdings and trading and price efficiency. Based on a sample of current constituents of the S&P 500 index and non-S&P 500 stocks with comparable size and turnover ratios, we find that prices become less efficient as indexed ownership grows, where price efficiency is measured by magnitude of post-earnings-announcement drift, deviations of price from the random walk, or return predictability from lagged order imbalances. It is robust to several liquidity measures and alternative specifications. We also examine the impact of indexed ownership and passive trading separately, and find that indexed investments affect price efficiency through both channels.

Our sample of passive institutional investors consists of 591 index funds, enhanced index funds and ETFs, and a number of closet indexers. The index and index-like funds are identified in several ways: keywords in fund names; 'activeness' of funds based on deviations from index compositions; and fit from regressions of fund returns on index returns. For subsequent analysis, we measure each stock's passive ownership as the percentage of shares held by any fund in our sample at the end of each quarter, and we measure passive trading volume as the sum of absolute holding changes over that quarter.

We measure price efficiency in several ways. First, we view post-earnings-announcement drift (PEAD) as an indicator of investors' underreaction to public information and use it as a framework to study price inefficiency. Next, we assume that an efficient stock price should follow a random walk and use deviation of stock prices from random walk as proxy of price inefficiency. Specifically, we adopt normalized Hasbrouck's (1993) intraday pricing error volatility, absolute value of first-order autocorrelation in daily returns, and absolute value of weekly to daily variance ratio as our empirical measures of price inefficiency. We believe that the intraday measures, compared to daily or even longer-horizon proxies, better capture deviation from efficient prices due to relatively quick adjustments of stock prices of S&P 500 firms within each trading day (Chordia et al., 2005). Meanwhile, the daily measures ensure that potential price inefficiency in longer horizons is not omitted. Finally, we assume that an efficient price should not be predicted by lagged order imbalance, thus we use return predictability from order imbalance over 5-minute intervals as a measure of price inefficiency.[4]

Our empirical tests use a sample period from 2002 to 2013, which is after the introduction of decimalization to avoid major institutional changes in price efficiency. We

[4] Chordia et al. (2005) find that lagged order imbalance has significant predictability on 30-minute future returns in 1996, on 15-minute future returns in 1991, and on 5-minute future returns in 2002. This reduction in return predictability probably reflects the gradually improved market efficiency over time. We choose a 5-minute interval to create our measure due to our post-2002 sample period.

first test the magnitude of post-earnings-announcement drift and find that PEAD increases along with passive ownership and, consistent with Bartov et al. (2000), decreases as (non-passive) institutional ownership increases. These findings are robust to alternative definitions of price drift, different drift windows, and alternative measures of earnings surprise. Following Fama and MacBeth (1973), we regress empirical measures of price efficiency on passive and non-passive institutional ownership. Consistent with our hypothesis, the deviation of both intraday and daily stock prices from a random walk, as well as the return predictability from past order imbalances, is positively correlated with passive ownership. We also reconfirm the inference of Boehmer and Kelley (2009) that (non-passive) institutional investors generally enhance price efficiency. We test for endogeneity that passive ownership increases for stocks that are characterized by price inefficiency, and confirm our main result with many alternative specifications.

In terms of an explanation for our results, further analysis reveals that information production, as measured by the number of analysts, is significantly and negatively associated with passive ownership but is positively related to non-passive institutional ownership. This evidence is consistent with the expectation that as indexed investors have no incentive for information acquisition and arbitrage, information production will be discouraged. Therefore, we believe that the reduced incentive for price discovery is primarily responsible for the negative impact of passive ownership on price efficiency. Moreover, passive investors may create limits to arbitrage through noise trader risk, which could deter or delay informed arbitrageurs from correcting the mispricing thus reducing price efficiency indirectly.

This study illustrates that while indexing is beneficial for passive investors, it exerts a negative externality by making it more difficult for stock prices to reflect information efficiently. As the fraction of passive ownership increases, prices are likely to become less efficient. Taken to an improbable and theoretical extreme of 100% indexing, no one would have an incentive to make prices informationally efficient. Of course, the current fraction of passive ownership is relatively small, thus its negative impact on market efficiency may not be striking or intuitive. However, given the likely rapid growth of indexing in the near future, it is reasonable to be cautious about its negative influence on market efficiency.

The rest of this essay is organized as follows. Section 2 describes the data sources and sample selection. Section 3 describes measures of price efficiency, liquidity, and other control variables. Section 4 presents our main results from post-earnings announcement drift and cross-sectional analyses, and corresponding robustness tests. Section 5 explores potential underlying mechanisms to explain our findings. Section 6 concludes.

2. Sample

2.1 Sample of Stocks

The primary empirical analysis in the essay is based on a stock sample that includes large and liquid U.S. common stocks. Specifically, at the end of each quarter from 2002 to 2013, we sort all the S&P 500 stocks into deciles by their market values estimated from the Fama and French (1992) approach and by their turnover ratios in the prior quarter, respectively. We include all the S&P 500 and non-S&P 500 stocks which have size larger than any stock in the size decile 1 (smallest) and turnover ratios greater than any stocks in the turnover decile 1 (lowest). This procedure generates a sample with, on average, 400 S&P 500 constituents and 194 non-S&P 500 stocks in each quarter. We require all stocks to have prices of at least $2 at the beginning of a quarter to avoid severe market microstructure issues. Since the focus of the essay is on price efficiency, we want to minimize the impact of external events. In particular, decimalization in 2001 has improved price efficiency considerably with an increase in liquidity that makes arbitrage less costly (Chordia et al., 2008). In order to examine changes in price efficiency due to passive ownership, we begin our sample in 2002 to avoid the effect of decimalization.

We choose the S&P 500 (and the comparable non-S&P 500 stocks) as our sample for several reasons. First, informational efficiency of stock prices is largely influenced by the information environment of the firm. Restricting our sample to only S&P 500 constituents and comparable non-S&P 500 firms generates a setting of similar informational environments across stocks, thus any inference about the association between indexed ownership and price efficiency will be more robust to potentially unobservable factors that may affect the informational environment of a firm. Second, the

S&P 500 index is the most popular index by assets indexed and one of the most popular indexes by the percent of assets indexed, which implies that S&P 500 stocks will exhibit a reasonable level of passive ownership. Meanwhile, the non-S&P 500 stocks typically have fairly low, if not zero, passive ownership. This combined sample thus provides sufficient cross-sectional dispersion in passive ownership for our empirical analysis.[5] Third, restricting the sample to S&P 500 constituents and comparable non-S&P 500 firms helps reduce the potential impact of infrequent trading, which could be severe in small, low-recognition firms. Finally, the S&P 500 index represents the U.S. equity market and any effect found among its components is likely to be important for the entire market.

Daily stock price, return, trading volume and shares outstanding are obtained from CRSP stock files. Intraday trade and quote data are obtained from NYSE Trade and Quote (TAQ) database. Information about earnings announcements are obtained from the Compustat. Analyst's earnings forecasts are obtained from the IBES. We obtain constituents of the S&P indices from the Compustat and of the Russell indices from Russell Investments. Total returns for the indexes are obtained from Bloomberg.

2.2 Sample of Passive Funds

Pure index funds and ETFs, which typically follow strictly passive investment strategies, are the major passive institutional investors on the market. However, we do not restrict our sample to pure index funds and ETFs, but also include enhanced index funds and closet indexers in order to construct a more complete measure of passive ownership. Although these funds may strategically adjust weights of some holdings based on their predictions about future price movements, they track indices passively and closely. Thus, their impact on price efficiency is closer to index funds than to their active peers.

Institutional holdings data from CRSP Mutual Fund database and Thomson Reuters (13f filings) are used to create the passive fund sample in four steps. First, we merge the

[5] A sample with only S&P 500 stocks has relatively low cross-sectional dispersion in passive ownership, as all S&P 500 index funds will hold each S&P stock in the same proportion thus the only cross-sectional dispersion comes from the presence of the S&P 500 stocks in other indexes.

CRSP mutual fund database, which provides indicators for index funds and ETFs, with the 13f database to identify index mutual funds and index ETFs. Second, we screen remaining funds for index-related funds on both CRSP Mutual Fund and 13f databases using keywords in their names. A fund is classified to be passive if it calls itself as an index fund, enhanced index fund, or an exchange traded fund (ETF).[6]

Third, we identify closet indexers in two ways. First, following Cremers and Petajisto (2009), we estimate the "Active Share" (*AS* hereafter) of each mutual fund from the 13f database. As suggested by Cremers and Petajisto (2009), any portfolio could be decomposed into a benchmark index portfolio plus a zero-net-value long-short portfolio. Thus, *AS* is a measure of the overall deviation of the weights of a fund's holdings from the benchmark index. For a pure index fund, *AS* will be close to zero, since the weight of each asset in the fund portfolio equals the asset weight in the benchmark index.[7] Second, we estimate a regression of daily fund returns over its entire life on corresponding benchmark index returns to obtain *R*-square. A fund with *R*-square close to one is more likely to follow passive strategies.[8] Since the benchmark indices for closet indexers are not explicitly stated, we tested ten indices for each fund and selected the lowest *AS* and the highest R-square for each fund. The indices are: S&P 500 index, S&P 500 Growth index, S&P 500 Value index, S&P 400 Mid-Cap index, S&P 600 Small-Cap index, S&P 100 index, Russell 1000 index, Russell 2000 index, NASDAQ 100 index, and the whole market portfolio obtained from CRSP stock files. To be included in the passive fund sample as a closet indexer, a fund quarter must have an *AS* less than 10% or an *R*-square above 99% in the prior year.

Finally, we exclude balanced funds, international funds, and bond funds from the sample.[9] The passive sample includes four types of passive institutional investors: 1) a

---

[6] Details of the keyword search are not reported here to save space.
[7] Theoretically, *AS* for a pure index fund with very low tracking error should be close to zero. However, as 13f reports generally ignore small holdings, the estimated *AS* will be higher than the actual value. The average estimate of *AS* over the life of an index fund in our sample could be as large as approximately 20%.
[8] We do not require the beta to be close to 1 because a passive fund may intentionally maintain a beta different from 1.00 by using leverage or by holding cash.
[9] Balanced funds, international funds, and bond funds are identified mainly by their names and country code provided in the 13f database. We also manually screen the sample to remove any of these funds. We check the fund prospectus for fund objective and strategies. Typically, a fund is removed from the sample when its investment strategy states that the manager actively chooses undervalued stocks.

total of 255 open-end equity index funds that aim to replicate the performance of a specific equity index by holding the index constituents in the same proportions as the index, 2) 47 enhanced index funds that reserve certain flexibility on position size and investment strategies, 3) 289 ETFs that track an index and are traded on stock exchanges,[10] and 4) a number of closet indexers. Appendix A reports the detailed procedure of the sample construction.

2.3 Trends in indexing

Table 1.1 shows the growth in indexing in the 2002-2013 period. The size of the U.S. equity market grew from $11 trillion in 2002 to over $26 trillion in 2013. In the same period, the value of passive funds almost quadruples from $309 billion to $1,865 billion. As a fraction of U.S. equity mutual funds and ETFs, passive funds increased from 12.80% in 2002 to 27.52% in 2013, and as a percent of the equity market from 2.80% to 7.09%. Passive ownership for S&P 500 stocks increased steadily from 4.54% in 2002 to 9.45% in 2013. While passive ownership has steadily increased, non-passive institutional ownership increased from 57.80% of U.S. equity in 2002 to 69.52% in 2007 before falling to 61.80% in 2013. The last column is the difference between 100% and holdings of institutional investors, which represents the sum of holdings of insiders, individuals, and any errors in reporting.

Please insert Table 1.1 here

3. Methodology

[10] This number is smaller than the actual number of U.S. equity ETFs for several reasons. First, we exclude ETFs that hold a significant percent of international equities. Second, ETFs that are reported jointly with index mutual funds are identified as index funds instead of ETFs in the sample. An example is Vanguard 500 Index Funds which has both investor shares and ETF shares. Third, the 13f database does not provide an ETF indicator, while the ETF indicators from CRSP are not able to identify all ETFs in the 13f database since the MFLINKS dataset does not provide a complete linkage between the CRSP mutual fund database and the 13f database.

3.1 Measure of relative informational efficiency of prices

Measures of price efficiency fall into several groups. First, price efficiency could be measured by returns to trading strategies based on market anomalies, such as short-term reversals, momentum, and post-earnings announcement drift. Second, under the assumption that efficient price follows a random walk process, deviations of stock prices from random walk could be a measure of relative inefficiency. This group of measures includes return autocorrelations, variance ratios, and the Hasbrouck (1993) pricing error variance. Third, the predictability of lagged order imbalances on future returns could indicate price inefficiency, as efficient price should not depend on previous order imbalance (Chordia et al., 2005). Fourth, without assumption about random walk, the relative price inefficiency could be inferred by the delay of stock return in response to market return (Hou and Moskowitz, 2005). Finally, price inefficiency could be measured by certain asset pricing models. For example, the Kalman filter estimation of mispricing used by Brennan and Wang (2010) is based on the assumption that the fundamental return follows an ex-post version of the Fama-French (1993) three factor model.

In this essay, we prefer the model-free assumption that efficient stock prices follow a random walk and is independent from lagged order imbalances, thus we measure price efficiency by the price's deviation from a random walk and its predictability by lagged order imbalances. Therefore, we adopt the pricing errors of Hasbrouck (1993), the daily first-order autocorrelation in stock returns, variance ratios of weekly returns to daily returns, and predictability of 5-minute stock returns by lagged 5-minute order imbalance as our measures of price inefficiency. We use pricing error as the principal measure, while the other three measures serve as robustness tests.

Compared with other measures of price inefficiency, Hasbrouck's (1993) pricing error has several advantages. First, Hasbrouck's measure is free from asset pricing model. Mispricing proxies that rely on specific asset pricing models, such as the Kalman filter in Brennan and Wang (2010), the price delay in Hou and Moskowitz (2005), or measures based on momentum profit or post-earnings announcement drift, may suffer from model misspecification. Second, as suggested by Boehmer and Kelley (2009), Hasbrouck's measure only captures price deviation from random walk caused by uninformed trading, so that deviations caused by informed trading will not be mistakenly measured as price inefficiency. Variance ratio and autocorrelation, however, do not distinguish between information-induced and non-information-induced deviations from random walk. Third, Hasbrouck's measure is estimated on a trade-to-trade basis, which captures most of the information involved in trading and places higher weight on periods with active information discovery. Variance ratio and autocorrelation, however, are measured with daily intervals leading to loss of intra-day information and place equal weights on periods with and without active price discovery. We expect that Hasbrouck's measure will provide more accurate and robust inferences compared to other empirical measures of price inefficiency.

The pricing error proposed by Hasbrouck (1993) measures the deviation between transaction prices and implicit efficient prices.[11] Specifically, the log transaction price, p_t, is defined as the efficient price, m_t, plus a transitory deviation, s_t:

$$p_t = m_t + s_t. \tag{1.1}$$

t indexes either transactions or natural time; m_t is defined as the expectation of the stock value given all available public information and is assumed to follow a random walk; s_t measures the deviation of transaction price from the efficient price. s_t is assumed to be a zero-mean covariance-stationary stochastic process with variance of σ_s^2, where σ_s^2 measures how closely the transaction price follows the efficient price. As σ_s^2 is associated with price volatility, we follow Boehmer and Kelley (2009) and several other studies[12] to normalize σ_s^2, denoted as *V*(s), by the variance of log transaction prices, *V*(p), to form a measure of relative price efficiency, *V*(s)/*V*(p). Ln[*V*(s)/*V*(p)] is used as principal metrics in the cross-sectional analysis.

Though the price adjustment process generally takes less than sixty minutes (Chordia et al., 2005) and should be well described by *V*(s)/*V*(p), we would like to capture

[11] A detailed analysis is in the Appendix B.

[12] Boehmer, Saar, and Yu (2005) use this measure to study the effect of the increased pre-trade transparency on stock price efficiency. Boehmer and Kelley (2009) use this measure to study the effect of institutional ownership on stock price efficiency. Boehmer and Wu (2012) use this measure to examine the relation between short selling and the price discovery process. Hotchkiss and Ronen (2002) use a variant of this measure, $MQ = 1 - 2\sigma_s^2/\sigma_p^2$, to examine the informational efficiency of corporate bond price.

potential price adjustment processes in longer horizons to enhance the robustness of our findings. The existence of long-horizon return anomalies, such as momentum, daily and weekly return autocorrelations, or post earning-announcement drift, indicate the existence of inefficient stock prices beyond each trading day. Hence we adopt two (inverse) efficiency measures based on daily and weekly returns.

They are the absolute value of first-order daily return autocorrelation, $|AC(1)|$ and the absolute value of weekly-to-daily return variance ratio, $|1-VR(1,5)|$. Both are associated with the magnitude of deviation of stock price from a random walk. Specifically, $|AC(1)|$ is estimated for each stock over each quarter by regressing daily returns on 1-day lagged returns, and $|1-VR(1,5)|$ is estimated for each stock over each quarter as the absolute deviation of the ratio of weekly return variance to (five times) daily return variance, where the weekly returns are calculated from a Wednesday to the next Tuesday to eliminate the weekend effect. All the daily and weekly returns are calculated by quote midpoints to eliminate potential bid-ask bounce.

We further adopt an additional measure based on intraday order imbalance. Chordia et al. (2005) find that order imbalances predict future returns over very short intervals (up to 5-minute in 2002), and they consider this a violation of strong-form efficiency. Following this concept, we consider greater predictability of returns by lagged order imbalance as an indicator of more severe price inefficiency. Specifically, we regress 5-minute quote-midpoint returns on the volume-based relative order imbalance over the previous 5-minute interval, and use the absolute value of the regression coefficient, $|\phi(OIB)|$, as an inverse measure of price efficiency.

Please insert Table 1.2 here

Table 1.2 reports descriptive statistics of the five efficiency measures over the period 2002 to 2013. The average $V(\mathrm{s})/V(\mathrm{p})$ of stocks in our sample is 0.007% and is decreasing over the sample period from 0.008% in early years to 0.004% in late years. $|AC(1)|$ and $|1-VR(1,5)|$ are relatively stable over time with average values of 0.106 and 0.286, respectively. $|\phi(OIB)|$, however, increases from 0.013% in early years to 0.018%

in later years. Overall, the level of market efficiency did not change materially over the observation period.

3.2 Measures of institutional ownership and trading

We define passive institutional ownership (*PO*) of a stock at the end of a quarter as the total shares held by any fund in the passive fund sample, scaled by total shares outstanding at the quarter end. Similarly, non-passive institutional ownership (*NPO*) is defined as the total shares held by any institutional investor (who files the 13f form) that does not belong to the passive fund sample.[13] Therefore, *PO* represents the fraction of shares held passively by institutional investors, while *NPO* represents the fraction of shares that are held by active institutional investors.

Since the 13f database contains only positions held, we are not able to precisely estimate trading volumes of either passive or non-passive institutional investors. Instead, we use changes in institutional holdings as a *lower bound* of institutional trading volume. Passive trading (*PT*) for each stock-quarter is estimated as the sum of absolute changes in passive holdings standardized by total shares outstanding, while non-passive trading (*NPT*) is estimated as the sum of absolute changes in non-passive holdings:

$$PT_{k,t} = \frac{\sum_{i=1}^{N_{passive}} \left|\Delta Holdings_{i,k,t}\right|}{ShareOut_{k,t}}, \tag{1.2}$$

$$NPT_{k,t} = \frac{\sum_{j=1}^{N_{non-passive}} \left|\Delta Holdings_{j,k,t}\right|}{ShareOut_{k,t}}. \tag{1.3}$$

Table 1.2 reports time-series average of quarterly cross-sectional means and standard deviations of ownership and trading variables. The average quarterly passive trading is 0.94% of total shares outstanding, which is equivalent to an annual turnover of

[13] Institutional holdings are obtained from the 13f database, while data on total shares outstanding is obtained from the CRSP stock files. For stock-quarters which report more institutional holdings than total shares outstanding, we set institutional ownership to 100% and calculate *PO* and *NPO* accordingly provided the institutional ownership in the previous or the following quarter is above 80%. Otherwise, we consider it to be an invalid observation.

3.76%. In contrast, the average quarterly non-passive institutional trading is 20.69% of total shares outstanding, or an annual turnover of 82.76%. The average *NPO* (69.56%) is about 16 times the average *PO* (5.65%), whereas average non-passive trading is about 12 times average passive trading. As expected, non-passive institutional investors trade much more than passive funds even after considering the difference in their ownership.

3.3 Measures in PEAD analysis

We use cumulative abnormal returns (*CAR*) estimated from the companion-portfolio approach as measures of PEAD. Specifically, the *CAR* for the window (t_1, t_2) is estimated as the daily abnormal return for firm *i* on date *t* as:

$$CAR_i(t_1, t_2) = \prod_{t=t_1}^{t_2}(1 + R_{i,t}) - \prod_{t=t_1}^{t_2}(1 + R_{p,t}) \quad (1.4)$$

where $R_{i,t}$ is the daily return of stock *i* and $R_{p,t}$ is the value-weighted daily return of a portfolio of stocks within the same size and book-to-market quintiles as stock *i*.[14]

Following Frazzini (2006), we use *CAR* (-2,1) – the cumulative abnormal return from 2 days preceding the announcement date to 1 day after – as the proxy for earnings surprise. Conventional measures of earnings surprise, such as the time-series model of actual earnings or consensus forecasts may not truly measure the market's expectation. As suggested by Frazzini (2006), however, *CAR* around the announcement date is the actual market reaction and does not rely on assumptions underlying any forecasts. Nevertheless, we also use two alternative measures of unexpected earnings in robustness tests. First, following Ayers, Li, and Yeung (2011), we define standardized unexpected earnings (SUE) as the seasonal difference in actual earnings standardized by the stock price at the end of the previous fiscal quarter:

[14] At the beginning of each quarter, we sort all stocks in our sample into quintiles by their size and book-to-market ratios, respectively. Size and B/M are estimated by the same approach of Fama and French (1992). We then construct 5×5 value-weighted benchmark portfolios, and use their returns to estimate *CAR*s of the stocks within the same size and B/M quintiles at that quarter.

$$SUE_{i,t} = \frac{EPS_{i,t} - EPS_{i,t-4}}{P_{i,t-1}}. \tag{1.5}$$

Second, we define analyst forecast error (*AFE*) as the difference between the actual EPS and the mean of the most recent analyst forecasts, scaled by the stock price at the end of the previous fiscal quarter.

Following Bartov et al. (2000) and Zhang (2012), among others, we use decile ranks of variables in our regression analysis. In each calendar quarter, each of the dependent and independent variables is ranked into deciles based on the cutoff values from previous quarter.[15]

3.4 Control variables

Chordia et al. (2008) find that liquidity stimulates arbitrage activities and enhances market efficiency. Four measures for liquidity (*ILLIQ*) are used in the cross-sectional analysis: 1) trade-weighted relative effective spread (*RES*), estimated as two times the absolute distance between actual transaction price and corresponding quote midpoint, scaled by the quote midpoint, and weighted by trade size, 2) time-weighted relative quoted spread (*RQS*), estimated as the absolute distance between bid and ask price, scaled by the quote midpoint, and weighted by time intervals between two quotes, 3) Amihud (2002) price impact measure of illiquidity (*Amihud*), estimated as change in rate of stock returns per million dollar trade and adjusted by equity market inflation, and 4) Liu (2006) no-trade-day measure of illiquidity, estimated as number of zero-trade days in a quarter and adjusted by turnover ratio in that quarter. *RES* is preferred since it measures the actual (relative) transaction costs for traders. However, we recognize that *RES* may underestimate illiquidity, since transactions are relatively infrequent during periods of low liquidity.

It is reasonable to expect that price efficiency is associated with stock size. Market value (*MV*) is measured as the number of shares outstanding multiplied by the closing price (CRSP items SHROUT and PRC) at the end of June of the previous year. We also control

[15] Each variable takes on a value of 0.05 for the smallest decile and 0.95 for the largest decile.

for dollar trading volume or turnover ratios, estimated as the dollar trading volume of a stock in a quarter scaled by total shares outstanding at the end of that quarter. Moreover, to control for potential effect of price discreteness on price inefficiency, we control for stock price at the end of the previous quarter. In addition, we use number of analysts following (*ANLY*) as a proxy for information production. *ANLY* is measured as the total number of analysts that report quarterly earnings forecasts for a stock (IBES item FPI = 6).

4. Empirical Results

4.1 Impact of indexed ownership on post-earnings-announcement drift (PEAD)

The following regression is estimated to study the relation between indexed ownership and PEAD:

$$DCAR_{i,t}(2,T) = \alpha_t + \gamma_1 DCAR_{i,t}(-2,1) + \gamma_2 DCAR_{i,t}(-2,1) * DPO_{i,t-1}$$
$$+\gamma_3 DCAR_{i,t}(-2,1) * DNPO_{i,t-1} + \gamma_4 DCAR_{i,t}(-2,1) * DMV_{i,t} + \gamma_5 DCAR_{i,t}(-2,1) * DPRC_{i,t}$$
$$+\gamma_6 DCAR_{i,t}(-2,1) * DRES_{i,t-1} + \gamma_7 DCAR_{i,t}(-2,1) * DVOL_{i,t} + \varepsilon_{i,t} \quad (1.6)$$

Quarterly observations of earnings announcement are pooled across our stock sample and over the period 2002-2013. We then regress cumulative abnormal returns on earnings surprises, an interaction term between earnings surprises and lagged indexed ownership (*DPO*), and an interaction term between earnings surprises and lagged non-passive institutional ownership (*DNPO*). We choose three windows, (2, 10), (2, 30), and (2, 60), to better capture price drifts over short and long periods after an earnings announcement. Control variables are constructed by interacting earnings surprises with the usual control variables like market value of the stock at the end of the previous calendar year (*DMV*), share price at the end of the previous year (*DPRC*), average relative effective spread over the previous quarter (*DRES*), and trading volume over the previous year (*DVOL*). Since quarterly decile ranks of variables are used, any potential time fixed-effects will automatically be controlled.

Please insert Table 1.3 here

Table 1.3 reports the main results. When earnings surprise is the only independent variable, *CAR*s over all the three drift windows are positively and significantly associated with earnings surprise, clearly indicating the existence of PEAD. When *PO* and *NPO* are included as independent variables, the positive relation between price drift after the earnings announcement and the earnings surprise is enhanced by *PO* while weakened by *NPO*. As the PEAD is generally considered an indication of investors' underreaction to earnings news, the positive impact of indexed ownership on PEAD is consistent with our hypothesis that indexing reduces stock price efficiency. The negative impact of non-passive institutional ownership on PEAD, on the other hand, is consistent with findings in the literature that (non-passive) institutional investors enhance stock price efficiency.

The above results, however, are unstable if raw values of each variable, instead of decile ranks, are used in the regression. Specifically, we try to use the average value in each decile to replace the decile ranks and find that the significant relations between the PEAD and the passive ownership, non-passive institutional ownership, and other control variables disappear or show conflict signs. This is probably due to the fact that the proxies of earnings surprise involve lots of noise. Nevertheless, the use of decile ranks has become a typical methodology in the literature of PEAD and it makes our results to be more comparable to other studies in the literature.

4.2 Alternative specifications on the analysis of post-earnings-announcement drift

We adopt several alternative specifications as robustness tests. First we address the robustness of our findings to the definition of earnings surprise by using two alternative proxies of earnings surprise – the standardized unexpected earnings (SUE) and standardized analyst forecast error (*AFE*). Both measures are widely used in the literature of PEAD, so our findings based these two measures will be easily comparable to previous studies. Panel A of Table 1.4 repeats our regression analysis but uses *AFE* to replace *CAR*(-2,1). When earnings surprise is the only independent variable, *CAR*s over the first two drift windows are positively and significantly associated with earnings surprise, though the relation is not statistically significant for the drift window (2, 60). When *PO* and *NPO* are included as independent variables, the relation between post-announcement *CAR* and *AFE* is still positive, and is statistically significant for drift window (2, 10) and (2, 60). *NPO*, in contrast, is negatively associated with post-announcement *CAR*. Panel B shows very similar results when *SUE* is used. Therefore, our finding that indexed ownership enhances PEAD is robustness to typical proxies of earnings surprise.

Please insert Table 1.4 here

Next, we address the robustness of our findings on the estimation of abnormal returns. Panel C of Table 1.4 reports the regression results when the abnormal return is defined as size-adjusted returns[16], while Panel D shows the results when raw returns, instead of abnormal returns, are used to measure returns around the announcement dates and the post-announcement price drifts. The results are still consistent with our hypothesis.

Another potential concern is the positive correlation in earnings surprise. Prior research, such as Bernard and Thomas (1990), has documented that earnings surprises show positive serial correlation for three quarters. Greater PEAD could thus be interpreted as either an indicator of delayed price response to earnings surprise or a result of forecasted positive earnings surprise in the subsequent quarter. If the second interpretation is true, magnitude of the PEAD may no longer be a valid proxy for price inefficiency. However, we believe that our regression results are not significantly affected by such a complication. If the magnitude of the PEAD in one quarter is increased by the accurate forecast of future positive earnings surprise, the magnitude of the PEAD in the following quarter(s) will be lowered. Overall, this may not significantly affect the regression coefficients.

4.3 Cross-sectional relation between indexed ownership and price efficiency

In addition to the PEAD analysis, we estimate a cross-section regression to directly investigate the relation between passive investments and stock price efficiency. We prefer

[16] At the beginning of each quarter, we sort all stocks in our sample into decile by their size, which is estimated by the same approach of Fama and French (1992). We then construct 10 value-weighted benchmark portfolios, and use their returns to estimate *CAR*s of the stocks within the same size decile at that quarter.

a cross-sectional approach to a time-series approach because, as passive funds are indexed and index constitution changes slowly, time-series variation in passive ownership of an index stock is fairly small. In contrast, cross-sectional dispersions in passive ownership and trading are much greater. First, the S&P 500 constituents in our sample typically have much higher passive ownership than non-S&P 500 stocks in the sample. Second, there is also some cross-sectional dispersion in passive ownership within the S&P 500 stocks, due to the presence of S&P 500 stocks in other indexes like the Russell and MSCI indexes and the existence of impure index funds such as closet indexers and enhanced indexers.

To formally examine the relation between indexed ownership and price efficiency, we estimate a multivariate cross-sectional regression following Fama and MacBeth (1973):

$$PE_{it} = \alpha_t + \beta_{PO,t}PO_{i,t-1} + \beta_{NPO,t}NPO_{i,t-1} + \gamma_{1,t}RES_{i,t-1} + \gamma_{2,t}MV_{i,t-1} + \gamma_{3,t}Price_{i,t-1} + \gamma_{4,t}TO_{i,t-1} + \varepsilon_{i,t}. \tag{1.7}$$

Specifically, one of the five price efficiency measures, PE_{it}, in each quarter is regressed on *PO*, *NPO*, and a lagged efficiency measure from the previous quarter, while controlling for illiquidity (*RES*), market value of the stock (*MV*), stock price (*Price*) as at the end of the previous quarter, and turnover ratio (*TO*). All independent variables are lagged by one quarter to prevent potential reverse causality and to reduce the potential impact of passive ownership on contemporaneous explanatory variables. For example, if passive ownership has an effect on both price efficiency and liquidity, using contemporaneous illiquidity ($RES_{i,t}$) as a control variable could lead to a downwardly biased coefficient on passive ownership ($\beta_{PO,t}$).[17] The lagged pricing efficiency measure is included to control for persistence in pricing efficiency over time and to control for the contemporaneous effect of $PE_{i,t-1}$ on $PO_{i,t-1}$ and $NPO_{i,t-1}$, if any. Many control variables are log transformed to control for non-normality. The cross-sectional regression is estimated in each quarter of our sample period, and the time-series mean of the quarterly coefficient estimates is used for inference. The standard errors are adjusted for residual autocorrelation and heteroskedasticity by the Newey and West (1987) approach. To make the coefficient

[17] Using contemporaneous variables leads to qualitatively and quantitatively similar regression results.

estimates in each quarter comparable across the entire sample period,[18] we follow Kumar (2009) to standardize all dependent and independent variables to have zero-mean and unit standard deviation on a quarterly basis.[19]

Please insert Table 1.5 here

Table 1.5 reports the main results from the cross-sectional analysis. Controlling for non-passive institutional ownership and other stock characteristics, each of the five inefficiency measures is positively and significantly related to passive ownership, indicating that indexed holdings are associated with greater deviation of stock price from random walk and stronger return predictability by lagged order imbalances. In contrast, non-passive institutional ownership is negatively and significantly related to four inefficiency measures, indicating their role in enhancing market efficiency. The results are consistent with Boehmer and Kelley (2009) that (active) institutional investors contribute to efficient stock prices.

4.4 Endogeneity

A positive association between inefficiency measures and indexed fund ownership may come from, though unlikely, a self-selection bias rather than causality. If passive funds prefer to hold stocks with lower price efficiency, a cross-sectional negative relation between indexed holding and price efficiency is expected even if passive holdings generate no impact on price efficiency. Though this problem could be significant in studies of active institutional investors, it is likely to be less serious when studying indexed institutional investors whose trading is mostly driven by investor flows or index changes rather than preference to stocks with certain characteristics. Nevertheless, we adopt two approaches

[18] Several dependent and independent variables suffer from heteroskedasticity over the sample period. There are noticeable increases in the cross-sectional standard deviations of *PO* and *PT* and decreases in the standard deviations of the normalized pricing error volatility, *V*(s)/*V*(p), and illiquidity measures RES, Amihud and Liu. Since the cross-sectional volatility of the dependent variables (i.e. *V*(s)/*V*(p)) are moving in the opposite direction compared to the volatility of the independent variables (i.e. *PO* and *PT*), coefficient estimates of *PO* and *PT* tend to have larger magnitudes in early years than in recent years, leading to the Fama-MacBeth regression results being more influenced by early years.

[19] The results are similar without variable standardization.

to preclude any possible self-selection bias. First, we exclude enhanced index funds and closet indexers from our passive fund sample, leaving only strictly passive index funds and ETFs, and repeat the cross-sectional regression. As presented in Panel A of Table 1.6, passive (strict index funds and ETFs) ownership is, again, associated with lower price efficiency. Next, we regress change in passive ownership over the current quarter on lagged change in price inefficiency measures, controlling for lagged changes in passive ownership, effective spread, stock capitalization, turnover ratio, and two dummy variables indicating S&P 500 additions and deletions (*SP_ADD* and *SP_DEL*):

$$\Delta PO_{it} = \alpha_t + \delta_t \Delta PE_{i,t-1} + \gamma_{1,t} \Delta PO_{i,t-1} + \gamma_{2,t} \Delta RES_{i,t-1} + \gamma_{3,t} \Delta MV_{i,t-1} + \gamma_{4,t} \Delta TO_{i,t-1} + \gamma_{5,t} SP\_ADD_{i,t} + \gamma_{6,t} SP\_DEL_{i,t} + \varepsilon_{i,t}. \qquad (1.8)$$

As reported in Panel B of Table 1.6, none of the five lagged inefficiency measure has a significantly positive relation with passive ownership. Therefore, there is no evidence of a self-selection bias.

Please insert Table 1.6 here

4.5 Alternative specifications of cross-section analysis

4.5.1 Alternative liquidity measures

As suggested by Chordia et al. (2008), liquidity stimulates arbitrage, which enhances market efficiency. Controlling for liquidity in our cross-sectional regression is important for two reasons. First, institutional investors, especially non-passive institutional investors, may have a preference towards liquid stocks due to their lower transaction costs. Thus, institutional holdings may be relatively efficiently priced simply because they are more liquid. Fortunately, this is not likely to be an important concern for indexed institutional investors who aim to track a specific stock index instead of actively looking for liquid stocks. Second, institutional trading and holdings may have an impact on liquidity, which in turn affects price efficiency. We estimate regressions similar to Table 1.5 but use alternative liquidity measures: relative quoted spread (*RQS*), the Amihud (2002) price impact measure (*Amihud*) based on daily price movements and trading volume, and the Liu (2006) no-trade-day measure of illiquidity (*Liu*). Panel A of Table 1.7 reports the coefficient estimates for *PO*, *NPO*, and the illiquidity measure in the three specifications. Consistent with Table 1.5, *PO* is positively and significantly associated with most of the five price inefficiency measures under alternative liquidity specifications, while *NPO* generally has negative coefficients. Therefore, our conclusion is robust to the use of alternative liquidity measures.

Please insert Table 1.7 here

4.5.2 Alternative specifications

For the first alternative specification, we reconsider inclusion of liquidity and turnover ratio as control variables. If passive ownership has a positive impact on liquidity, controlling for liquidity and turnover ratio could make its negative impact on price efficiency look more pronounced. Second, if S&P 500 index constituents generally have lower price efficiency for reasons other than passive funds, the negative relation between passive ownership and price efficiency could just be a coincidence without implying causality. To eliminate such a possibility, we add an additional control variable which equals to 1 when a stock is a member of the S&P 500 index at that quarter and zero otherwise. Third, following Boehmer and Kelley (2009), we add the lagged dependent variable as an additional control variable, so that our results will be more comparable with Boehmer and Kelley (2009). Fourth, we use contemporaneous independent variables instead of lagged ones. Finally, instead of using the $2 filter, we remove stocks with a stock price of less than $5 at the beginning of a quarter to further eliminate potential microstructure biases. As presented in Panel B and C of Table 1.7, each of the above specifications generates qualitatively and quantitatively similar results as Table 1.5. Moreover, using share trading volume or dollar trading volume to replace turnover ratio or using raw variables instead of standardized variables leads to similar results.

5. Indexing and Price Efficiency - Potential Explanations

We consider four explanations for our results. First, higher indexed ownership possibly implies a reduced incentive for indexed investors to acquire information. Consequently, it may reduce the production of information and raise the cost of informed arbitrage for active investors. Second, as suggested by De Long et al. (1990), informed investors may not be willing to take risky positions to correct mispricing due to the existence of noise trader risk. The presence of passive investors, who may create noise trader risk, may reduce the incentive and effectiveness of informed investors' effort of price discovery, leading to more persistent price inefficiency. Third, indexed ownership may serve as a proxy for passive trading, which is mostly uninformed. In addition, index-related trading due to index changes is unidirectional, which may cause prices to move away from fundamentals (Harris and Gurel, 1986; Chen, Noronha, and Singal, 2004). Finally, with more uninformed investors choosing index funds and trade infrequently, the proportion of informed trades in the market will increase and the trading volume and frequency may decrease. Consequently, cost of trading could increase due to wider quote spreads and lower turnover, making prices less efficient.

5.1 Reduced information acquisition and production

Indexed or passive ownership represents shareholders who have no desire to acquire information or to incorporate that information in stock prices. These investors rely on index providers like Standard and Poor's, Russell Investments, and MSCI to make changes to an index as necessary. Index funds execute those changes in their portfolios. The passive investors are interested in matching the index, not in earning abnormal returns. In addition, holding a basket of securities could reduce the incentive, or say necessity, of informed arbitrage because random mispricing in index stocks is likely to cancel out: lower returns from overpriced stocks are set off against higher returns from underpriced stocks.

With a focus among passive investors on reducing the cost of acquiring information and the absence of motivation to arbitrage could reduce the aggregate demand for new information. The reduction in demand for new information would reduce the production of information. While active investors may continue to demand and produce new information, information production will become more costly as fewer investors pay for that information. Thus, we expect a negative relation between passive ownership and information production. In Table 1.8, using number of analysts following a stock as the proxy for information production, we regress the number of analysts against passive ownership and non-passive ownership based on all stocks in our sample:

$$ANLY_{i,t} = \alpha_t + \beta_{PO,t}PO_{i,t-1} + \beta_{NPO,t}NPO_{i,t-1} + \delta_t ANLY_{i,t-1} + \gamma_{1,t}MV_{i,t}$$
$$+\gamma_{2,t}TO_{i,t} + \gamma_{3,t}SP500D + \gamma_{4,t}R1000D + \varepsilon_{i,t}, \qquad (1.9)$$

where *MV* is the market value of the stock estimated by the approach of Fama and French (1992), *TO* is the turnover ratio over the previous year. As index additions are associated with a significant increase in number of analysts (Yu, 2008) as well as a significant increase in passive ownership, controlling for index membership is crucial to eliminate any false positive relation between passive ownership and number of analysts. Therefore, we add two dummy variables indicating current memberships in the S&P 500 index (*SP500D*) or the Russell 1000 index (*R1000D*).

Consistent with our conjecture, number of analysts is negatively related with passive ownership in all the three regression specifications in Table 1.8, implying that less information is being produced for firms with greater fractions of passive investors[20]. Moreover, consistent with Brennan and Subrahmanyam (1995), number of analysts is positively related with non-passive ownership, indicating that non-passive institutions indirectly facilitate information production.

Please insert Table 1.8 here

5.2 Lowered incentive of arbitrage

[20] Yu (2008) finds that analyst coverage is negatively associated with earnings management because, in addition to their role in information production, analysts serve as external monitors. Therefore, lower analyst coverage induced by indexing may encourage the firms to intentionally create information asymmetry by earnings management, resulting lower price efficiency.

A potential challenge to the above argument comes from the existence of informed traders. Even though ordinary investors may face higher cost of information acquisition caused by indexed investors, it is still likely that a small group of investors have relatively easy access to superior information. Thus, the negative impact of indexed investors on price efficiency could be largely offset by arbitrage activities of such informed traders. However, as suggested by De Long et al. (1990), informed investors may not be willing to take risky positions against noise traders (indexed investors), due to the existence of noise trader risk – the mispricing caused by noise traders could become even more extreme before its disappearance. The presence of passive investors, therefore, may reduce the incentive and effectiveness of informed investors' effort of price discovery. For example, Morck and Yang (2001) suggest that the increasing demands for S&P 500 stocks from passive funds induce ever-increasing overpricing in these stocks. If this finding is true, then short-selling will not be a rational strategy even if the informed investors clearly recognize the existence of overpricing, as the increasing demand from indexed investors will push the index stocks to be even more overpriced instead of converging to fundamental values, at least in the short term.

5.3 Consequences of passive trading

Index fund managers trade based on fund flows, which is usually uninformed about individual stocks or industries. They also trade around index changes, which is not only uninformed, but can also cause herding among passive investors and temporarily move prices away from fundamentals (Harris and Gurel, 1986; Lynch and Mendenhall, 1997; Chen, Noronha, and Singal, 2004). Research on market microstructure suggests that price efficiency will decline in the presence of such uninformed trading (Glosten and Milgrom, 1985; Kyle, 1985). On the other hand, passive trading has fairly small volumes due to their passive nature. As reported in Table 1.2, average passive trading volume is only 4.5% of average non-passive trading volume. Therefore, whether passive trading could generate a significant negative impact on price efficiency becomes an empirical question.

To evaluate the effect of passive ownership on price efficiency, PE, we examine the relation between passive trading on price efficiency using the following cross-sectional regression:

$$PE_{it} = \alpha_t + \beta_{PT,t}PT_{i,t} + \beta_{NPT,t}NPT_{i,t} + \beta_{PO,t}PO_{i,t-1} + \beta_{NPO,t}NPO_{i,t-1} + \gamma_{1,t}RES_{i,t-1} + \gamma_{2,t}MV_{i,t-1} + \gamma_{3,t}PRC_{i,t} + \gamma_{4,t}TO_{i,t-1} + \varepsilon_{i,t}. \quad (1.10)$$

If the negative relation between passive ownership and price efficiency is caused by passive trading, the coefficient on $PT_{i,t}$ in equation (10) should be significantly positive. We use contemporaneous trading instead of lagged trading because potential impacts of trading on stock prices should take place immediate. As mentioned earlier, a self-selection bias is not likely to exist in the trading of passive institutions and we actually find no evidence to support such a bias.

Please insert Table 1.9 here

The results in Table 1.9 suggest no significant impact of *PT* on price efficiency. Among the give measures of price inefficiency, only the Hasbrouck's (1993) measures has a positive coefficient that is marginally significant, while other four measures are not significantly related to *PT*. In addition, *PO* is positively related to all the five price inefficiency measures even after controlling for *PT*. These results suggest that passive trading seems not to be a source of indexed investors' negative impact on price efficiency. Of course, our measure of passive trading, which is based on absolute changes in holdings, is the lower bound of the actual trading volume and is likely to underestimate the actual trading volume. More precise data on institutional trading volume, if available, may deliver more reliable results.

5.4 Effect on liquidity

The negative effect of passive ownership on stock liquidity could raise the cost of informed arbitrage, which in turn impairs price efficiency. Market makers compensate their loss from trading with informed traders by the profits from trading with liquidity traders (Kyle, 1985). As more uninformed traders are attracted to index funds with less trading, a higher proportion of traders will be informed traders. Realizing the increase in informed traders, market makers may widen spreads thereby increasing the cost of trading. Moreover, as indexed investors trade much less than their active peers, higher fractions of indexed investors may imply lower turnover and trading frequencies. If indexed ownership or trading decreases liquidity, it may weaken price efficiency, as lower liquidity raises transaction costs and discourages informed arbitrage. We estimate equation (1.11) as a Fama-MacBeth cross-sectional regression of stock illiquidity and the results are reported in Table 1.10.

$$ILLIQ_{it} = \alpha_t + \beta_{PO,t}PO_{i,t-1} + \beta_{NPO,t}NPO_{i,t-1} + \gamma_{2,t}MV_{i,t-1} + \gamma_{3,t}PRC_{i,t} + \gamma_{4,t}SP500Dummy + \varepsilon_{i,t}. \quad (1.11)$$

Please insert Table 1.10 here

There are three results of interest. First, passive ownership has very limited impact on liquidity. As reported in Panel A, higher passive ownership is associated with lower effective spreads, but is not significantly related to quoted spread, price impact, no-trade days, or turnover. Second, Panel B shows that passive trading does not have significant impact on any of the five liquidity (illiquidity) measures. Third, trading from non-passive institutions greatly enhance liquidity, as the quoted spread, price impact, and no-trade days are significantly reduced and turnover is significantly increased by higher *NPT*. In sum, we do not find any evidence supporting a negative impact of passive holdings and trading on stock liquidity. Therefore, liquidity is not likely to be a channel of the negative effect of indexing on price efficiency.

6. Conclusion

Index funds and indexed investing have been promoted by academics and practitioners over the last 50 years as an inexpensive and effective way to hold a diversified portfolio. As a result, the indexed investment sector has grown and today accounts for more than 7% of the total equity market and over 23% of the U.S. equity mutual fund sector.

While advantages of index investing are significant, there are negative externalities that passive investors impose on other market participants and the economy by making stock prices less efficient. In a sense, index investors are free riders on rest of the market: active traders produce information and trade to earn abnormal returns. In the process, they contribute to market efficiency. Index investors, on the other hand, use these efficient prices to invest but without directly contributing to making those prices efficient. Their trades are primarily liquidity-driven, information-less trades motivated either by index changes or by investor flows.

Consistent with the above notion and based on a sample of large and liquid U.S. stocks over the period 2002 to 2013, we find that indexing reduces informational efficiency of stock prices, and stocks with a higher level of indexing, as measured by passive ownership, have less informative prices. We examine explanations for the decrease in price efficiency. We find that the relation is not explained by persistence in price efficiency, size, or endogeneity. The relationship is robust to several intraday and daily price efficiency measures and alternate liquidity measures. We distinguish between the effects of indexed ownership and passive trading on price efficiency, and find that indexing affects price efficiency probably through reduced information acquisition and arbitrage. We also argue that indexed investor may indirectly lower price efficiency by reducing informed investors' incentive to arbitrage.

Appendix 1.A: Selection of Index Funds and ETFs

In the first step, we pick up all funds that are classified as either an index fund or an ETF by the CRSP index fund and ETF indicators. To identify potential index funds and ETFs that are not marked by the indicators, we then screen fund names in the 13f database by keywords. For index funds, we look for the following keywords: 'INDEX', 'IND', 'IDX', INDE', 'S&P 500 I', 'S&P 500I', 'S&P 400 I', 'S&P 400I', 'S&P 600 I', 'S&P 600I', 'S&P500IND', 'S&P400IND', 'S&P600IND', 'RUSSELL 1000', 'RUSSELL 2000', 'RUSSELL 3000', and 'VANGUARD'. For ETFs, we look for the following keywords: 'EXCHANGE TRADED', 'EXCHANGE-TRADED', 'ETF', 'ISHARES', 'POWERSHARES', 'PROFUNDS', 'SPDR S&P', 'SPDR DOW', 'SPDR DJ, 'RYDEX', 'SPA MG', 'MARKET GRADER', and 'QQQ'.

To exclude bond funds, balanced funds, and funds that hold substantially derivatives from our sample, we remove funds with the following keywords in their names: 'BOND', 'INFLATION', 'TREASURY', ' BD ', 'LEHMAN', 'BARCLAY', 'OPTION', 'HEDGE', 'BALANCE', 'ALLOC', 'ASSET AL', 'MULTI ASSET', and 'PRINCIPAL PROTECTION'.

To exclude international funds, we require that the country code in 13f is either blank or 'UNITED STATES'. Further, we remove funds with the following keywords in their names: 'EURO', 'FRANCE', 'GERMAN', 'CANADA', 'CANADIAN', ' HK ', 'JAPAN', ' SING', 'INDA', 'INDU', 'INDI', 'INDO', 'NETH', 'SWITZ', 'ITALY', 'SPAIN', 'ASIA', ' GLOBAL', ' NIKKEI', 'FT-SE', 'FTSE', ' EM ', ' EMER ', ' BRIC', ' EUR', ' UK ', ' INT', 'AUSTRLA', ' JAP', 'CNDN', ' CDN', 'PACIF', ' TRU ', 'LATIN', ' EMER', ' EMG', 'EMRG', 'LAT AME', 'KINDOM', 'CHILE', ' JPN', 'TURKEY', 'DEVELOPE', 'ENERGY', 'BRAZIL', 'KOREA', 'BELG', 'MALAYSIA', 'SWEDEN', 'AUSTRIA', ' EMU', 'SOUTH AFR', 'TAIWAN', 'INDONESIA', 'STOXX', 'THAI', 'EX US', 'INDEKS', 'NIKKO', 'TOKYO', 'HANG SENG', 'JPA', 'SIMCAV', 'TOPIX', 'EAFE', 'SPHINX', 'WARBURG', 'FOND', 'TSX', 'AMER EXEMPT', 'TSE', 'GOLDEN DRAGO', 'AVENIR ALIZES', 'FINORD INDEX AMERIQUE', and 'ASX'.

Finally, we manually check the investment objective and strategies of each fund from its prospectus and remove funds that are not passive equity funds. We remove funds with the following ID number in the 13f database: 526, 583, 697, 787, 792, 1366, 1469, 1588, 1884, 2231, 2373, 2468, 2518, 2637, 2676, 2875, 2882, 2887, 2965, 3300, 3300, 5040, 7679, 12065, 12065, 12096, 12707, 12760, 12877, 13000, 13143, 13235, 13256, 14266, 14499, 16561, 16570, 16598, 18009, 18252, 20075, 21002, 21888, 22461, 22616, 23300, 23645, 26775, 28900, 28908, 29093, 34560, 36077, 36578, 36593, 45638, 47191, 47224, 47959, 48003, 48160, 49335, 51143, 51527, 51652, 51894, 53700, 53705, 53800, 53900, 53933, 54440, 55633, 56500, 58099, 58852, 60100, 61423, 63079, 64362, 64635, 64635, 64803, 64804, 64805, 64816, 64960, 66970, 67996, 68391, 68392, 70032, 71917, 72523, 72986, 73268, 73290, 73424, 73695, 73695, 74147, 74285, 75703, 75704, 75708, 76021, 76021, 76734, 77497, 77498, 77889, 77941, 78219, 78580, 79882, 80729, 80730, 80811, 80857, 80859, 81110, 81200, 83285, and 83380.

Appendix 1.B: Estimation of Hasbrouck's (1993) Pricing Error

Intraday trade and quote data obtained from the NYSE TAQ database are used for estimation of σ_s. Following Boehmer and Kelly (2009), we use quotes and trades that are within the regular trading hours (9:30AM-4:00PM) and ignore overnight price changes. A quote is removed if the ask price is lower than the bid price, if the bid price is lower than \$0.10, or if the bid-ask spread is higher than 25% of the quote midpoint. To be eligible for estimation, a trade is required to have a value of zero in TAQ's CORR field, marked as '*', '@', '@F', 'F', 'B', 'E', 'J', 'K', or blank in TAQ's COND field, and have a positive trade size and price. A trade is removed if its price differs by more than 30% from the previous trade. We ignore the natural times but view transactions as untimed sequences. This approach is preferable since it gives more weights to periods with heavier price discovery activities, represented by more transactions, and uses information delivered from every single transaction. Following Hasbrouck (1993), we estimate the lower bound for σ_s using a vector autoregression (VAR) model with five lags over the four-variable set $\boldsymbol{X}_t = \{r_t, \boldsymbol{x}_t\}'$, where $r_t = p_t - p_{t-1}$ and $\boldsymbol{x}_t$ is a 3×1 vector of the following trade variables: 1) sign of trading direction that takes value of 1 if the transaction is buyer-initiated value of -1 if it is seller-initiated, and value of 0 for a quote midpoint transaction, 2) signed trading volume, and 3) the signed square root of trading volume. Following Harris (1989) and Lee and Ready (1991), we classify a trade as buyer-initiated (seller-initiated) if the transaction price is above (below) the prevailing quote midpoint. The inclusion of square root of trading volume aims to allow for concave dependencies in both m_t and s_t. In each month, we estimate *V(s)* for stocks that have at least 100 trades over that month. Specifically, the joint process of $\boldsymbol{X}_t$ is described by a five-lag VAR model:

$$\boldsymbol{X}_t = \boldsymbol{B}_1\boldsymbol{X}_{t-1} + \boldsymbol{B}_2\boldsymbol{X}_{t-2} + \boldsymbol{B}_3\boldsymbol{X}_{t-3} + \boldsymbol{B}_4\boldsymbol{X}_{t-4} + \boldsymbol{B}_5\boldsymbol{X}_{t-5} + \boldsymbol{u}_t, \quad (1.B.1)$$

where $\boldsymbol{B}_k$ is the 4×4 coefficient matrix for lag *k*; $\boldsymbol{u}_t$ is a 1×4 vector of zero-mean error terms with $E(u_{i,t}, u_{j,t}) = 0$. The VAR model is then transformed into a five-lag approximation of vector moving average (VMA) representation:[21]

$$\boldsymbol{X}_t = \boldsymbol{u}_t + \boldsymbol{A}_1\boldsymbol{u}_{t-1} + \boldsymbol{A}_2\boldsymbol{u}_{t-2} + \boldsymbol{A}_3\boldsymbol{u}_{t-3} + \boldsymbol{A}_4\boldsymbol{u}_{t-4} + \boldsymbol{A}_5\boldsymbol{u}_{t-5}. \quad (1.B.2)$$

Variance of pricing error is expressed by:

$$\sigma_s^2 = \sum_{j=0}^{4} [\gamma_{1,j} \;\; \gamma_{2,j} \;\; \gamma_{3,j} \;\; \gamma_{4,j}]\, Cov(\boldsymbol{u}) [\gamma_{1,j} \;\; \gamma_{2,j} \;\; \gamma_{3,j} \;\; \gamma_{4,j}]', \quad (1.B.3)$$

where

$$\gamma_{i,j} = -\sum_{k=j+1}^{5} A_{k,1,i}, \quad (1.B.4)$$

and $Cov(\boldsymbol{u})$ is the residual covariance matrix from the VAR model. We use the average of the monthly estimates of σ_s^2 in a quarter as the pricing error variance of that quarter.

[21] VMA lags beyond five are assumed to have little impact and are ignored to simplify the estimation process. As mentioned by Hasbrouck (1993), more lags generally increase the estimates of pricing error variance but make little change in the cross-sectional ranking.

References 1

Acharya, Viral V., and Lasse Heje Pedersen, 2005. Asset pricing with liquidity risk. Journal of Financial Economics 77, 375-410.

Amihud, Yahov, 2002. Illiquidity and stock returns: Cross-section and time-series effects. Journal of Financial Markets 5, 31-56.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006. The cross-section of volatility and expected returns. Journal of Finance Vol. 61, Issue 1, 259-299.

Ayers, Benjamin C., Oliver Zhen Li, and P. Eric Yeung, 2011. Investor Trading and the Post-Earnings-Announcement Drift. Accounting Review Vol. 86, No. 2, 385-416.

Bartov, Eli, Suresh Radhakrishnan, and Itzhak Krinsky, 2000. Investor Sophistication and Patterns in Stock Returns after Earnings Announcements. Accounting Review Vol. 75, No. 1, 43-63.

Basak, Suleyman, and Anna Pavlova, 2013. Asset Prices and Institutional Investors. American Economic Review 103(5): 1728-1758.

Boehmer, Ekkehart, and Eric K. Kelley, 2009. Institutional investors and the informational efficiency of prices. Review of Financial Studies 22 n 9, 3563-3594.

Boehmer, Ekkehart, Gideon Saar, and Lei Yu, 2005. Lifting the Veil: An analysis of pre-trade transparency at the NYSE. Journal of Finance Vol. 60, Issue 2, 783-815.

Boehmer, Ekkehart and Juan Wu, 2012. Short selling and the price discovery process. Review of Financial Studies Vol. 26, No. 2, 287-322.

Brennan, Michael J. and Ashley W. Wang, 2010. The mispricing return premium. Review of Financial Studies Vol. 23, No. 9, 3437-3468.

Brennan, Michael J. and Avanidhar Subrahmanyam, 1995. Investment analysis and price formation in securities markets. Journal of Financial Economics Vol. 38, Issue 3 361–381.

Chan, Louis K. C., and Josef Lakonishok, 1993. Institutional trades and intraday stock price behavior. Journal of Financial Economics Vol. 33, Issue 2, 173-199.

Chen, Honghui, Gregory Noronha, and Vijay Singal, 2004. The price response to S&P 500 index additions and deletions: Evidence of asymmetry and a new explanation. Journal of Finance Vol. 59, Issue 4, 1901-1930.

Chiyachantana, Chiraphol N., Pankaj K. Jain, Christine X. Jiang, Robert Wood, 2004. International evidence on institutional trading behavior and price impact. Journal of Finance Vol. 59, Issue 2, 869-898.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2005. Evidence on the speed of convergence to market efficiency. Journal of Financial Economics Vol. 76, Issue 2, 271–292.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2008. Liquidity and market efficiency. Journal of Financial Economics 87, 249-268.

Cremers, Martijn, and Antti Petajisto, 2009. How active is your fund manager? A new measure that predicts performance. Review of Financial Studies 22 n 9, 3329-3365.

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann, 1990. Noise trader risk in financial markets. Journal of Political Economy Vol. 98, No. 4 703-738

Fama, Eugene F., 1991. Efficient Capital Markets: II. The Journal of Finance Vol. 46, No. 5, 1575-1617.

Fama, Eugene F. and Kenneth R. French, 1992. The cross-section of expected stock returns. Journal of Finance Vol. 47, Issue 2, 427-465.

Fama, Eugene F. and Kenneth R. French, 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics Vol. 33, Issue 1, 3-56.

Fama, Eugene F. and James D. MacBeth, 1973. Risk, return, and equilibrium: empirical tests. Journal of Political Economy Vol. 81, No. 3, 607-636.

Frazzini, Andrea, 2006. The Disposition Effect and Underreaction to News. Journal of Finance Vol. 61, No. 4, 2017–2046.

Galbraith, John W., Aman Ullah, and Victoria Zinde-Walsh, 2002. Estimation of the vector moving average model by vector autoregression. Working paper.

Glosten, Lawrence R., and Paul R. Milgrom, 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics Vol. 14, Issue 1, 71-100.

Goetzmann, William N., and Massimo Massa, 2003. Index funds and stock market growth. The Journal of Business Vol. 76, No. 1, 1-28.

Greenwood, Robin and David Thesmar, 2011. Stock price fragility. Journal of Financial Economics 102(3), 471-490.

Griffin, John M., Jeffrey H. Harris, Selim Topaloglu, 2003. The dynamics of institutional and individual trading. Journal of Finance Vol. 58, Issue 6, 2285-2320.

Grossman, Sanford J., and Joseph E. Stiglitz, 1980. On the impossibility of informationally efficient markets. American Economic Review, Vol. 70, No. 3, 393-408.

Harris, Lawrence, 1989. A day-end transaction price anomaly. Journal of Financial and Quantitative Analysis 24, 29-45.

Harris, Lawrence, and Eitan Gurel, 1986. Price and volume effects associated with changes in the S&P 500 list: New evidence for the existence of price pressures. Journal of Finance Vol. 41 No. 4, 815-829.

Hasbrouck, Joel, 1988. Trades, quotes, inventories and information. Journal of Financial Economics 22, 229-252.

Hasbrouck, Joel, 1993. Assessing the quality of a security market: A new approach to transaction-cost measurement. Review of Financial Studies Vol. 6, No. 1, 191-212.

Hegde, Shantaram P. and John B. McDermott, 2003. The liquidity effects of revisions to the S&P 500 index: an empirical analysis. Journal of Financial Markets Vol. 6, Issue 3 413-459.

Hotchkiss, Edith S. and Tavy Ronen, 2002. The informational efficiency of the corporate bond market: An intraday analysis. Review of Financial Studies Vol. 15, No. 5, 1325-1354.

Hou, Kewei and Tobias J. Moskowitz, 2005. Market frictions, price delay, and the cross-section of expected return. Review of Financial Studies Vol. 18, No. 3, 981-1020.

Jones, Charles M., and Marc L. Lipson, 1999. Execution costs of institutional equity orders. Journal of Financial Intermediation Vol. 8, Issue 3, 123–140.

Jones, Charles M., and Marc L. Lipson, 2001. Sixteenths: direct evidence on institutional execution costs. Journal of Financial Economics Vol. 59, Issue 2, 253–278.

Keim, Donald B., and Ananth Madhavan, 1997. Transactions costs and investment style: an inter-exchange analysis of institutional equity trades. Journal of Financial Economics Vol. 46, Issue 3, 265–292.

Kelly, Patrick J., 2007. Information efficiency and firm-specific return variation. Working Paper, New Economic School, Moscow.

Kumar, Alok, 2009. Who gambles in the stock market? The Journal of Finance Vol. 64 No.4, 1889-1933.

Kyle, Albert S., 1985. Continuous auctions and insider trading. Econometrica Vol. 53, No. 6, 1315-1335.

Lee, Charles M. C., and Mark J. Ready, 1991. Inferring trade direction from intraday data. Journal of Finance Vol. 46, Issue 2, 733-746.

Liu, Weimin, 2006. A liquidity-augmented capital asset pricing model. Journal of Financial Economics, Vol. 82, Issue 3, 631-671.

Lo, Andrew W., and A. Craig MacKinlay, 1988. Stock market prices do not follow random walks: evidence from a simple specification test. Review of Financial Studies Vol. 1, Issue 1, 41-66.

Lynch, Anthony W., and Richard R. Mendenhall, 1997. New evidence on stock price effects associated with changes in the S&P 500 index. Journal of Business 70, 351–383.

Morck, Randall and Fan Yang, 2001. The mysterious growing value of S&P 500 membership. Unpublished working paper.

Newey, Whitney K., and Kenneth D. West, 1987. A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. Econometrica Vol. 55, No. 3, 703-708.

Shu, Tao, 2007. Institutional trading and stock price efficiency. Ph.D. dissertation. University of Texas at Austin.

Wurgler, Jeffrey, 2011. On the Economic Consequences of Index-linked Investing, In “Challenges to Business in the Twenty-First Century”, Chapter 3

Wurgler, Jeffrey, and Ekaterina Zhuravskaya, 2002. Does arbitrage flatten demand curves for stocks? Journal of Business, Vol. 75, No. 4, 583-608.

Yu, Fang, 2008. Analyst coverage and earnings management. Journal of Financial Economics Vol. 8, Issue 2 245-271.

Zhang, Li, 2012. The Effect of Ex Ante Management Forecast Accuracy on the Post-Earnings-Announcement Drift. Accounting Review Vol. 87, No. 5, 1791-1818.

Table 1.1: Summary statistics of passive fund sample

The sample of passive funds includes a total of 591 U.S. equity index funds, enhanced index funds, ETFs, and 'closet indexers' over the sample period of 2002 to 2013. Values of passive funds and institutional investors are estimated from their reported (in 13-f files) holdings and the corresponding stock price. Institutional investors include all institutions that file quarterly 13-f reports. Column [2] presents the total capitalization of the U.S. equity market. Column [3] shows the total capitalization of U.S. equity mutual funds and ETFs. Column [4] reports the total number of funds in the sample in each year. Column [5] reports the total value of passive fund holdings. Column [6] reports the total value of holdings of U.S. institutional investors who file the 13f form. Column [7] reports the market share of the passive fund sample, measured as total passive fund holdings divided by total U.S. equity market capitalization. Column [8] reports the fraction of market capitalization of all passive funds in the total capitalization of the U.S. equity market. Column [9] presents the percentage of passive funds in U.S. mutual fund and ETF sector. Column [10] reports the non-passive institutional share of the U.S. equity market. Column 11 contains the percent equity in the hands of non-institutional holders like insiders and individuals.

Year	Total Cap of U.S. Equity Market ($ billions)	Size of U.S. Equity Mutual Funds and ETFs ($ billions)	NO. of Passive Funds	Total Value of Passive Funds ($ billions)	Total Value of Inst. Investors ($ billions)	% PO of S&P 500 Stocks	% Market Cap of Passive Funds	Fraction of Passive Funds in U.S. Equity Mutual Funds and ETFs	% Market Cap of Non-Passive Inst. Investors	Other Investors as a % of Market Cap
Source	CRSP	Investment Company Institute	Thomson Reuters and CRSP Mutual Funds		13f filings	Calculated	[5]/[2]	[5]/[3]	[6]/[2] minus [8]	1 minus [8] & [10]
[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]
2002	11,026.99	2,414.81	260	309.11	6,682.16	4.54%	2.80%	12.80%	57.80%	39.40%
2003	14,577.77	3,320.30	261	443.34	9,190.98	4.88%	3.04%	13.35%	60.01%	36.95%
2004	16,449.40	3,906.15	279	586.04	10,603.06	5.08%	3.56%	15.00%	60.90%	35.54%
2005	17,369.92	4,260.92	279	614.94	11,838.86	5.00%	3.54%	14.43%	64.62%	31.84%
2006	19,599.79	4,890.89	271	735.51	13,810.55	5.20%	3.75%	15.04%	66.71%	29.54%
2007	20,190.51	5,204.80	426	830.09	14,867.53	6.03%	4.11%	15.95%	69.52%	26.36%
2008	12,128.88	3,135.86	461	619.45	8,496.12	8.58%	5.11%	19.75%	64.94%	29.95%
2009	15,804.56	4,033.89	432	822.73	10,939.38	7.86%	5.21%	20.40%	64.01%	30.78%
2010	18,490.48	4,626.20	416	1,007.63	12,770.74	8.29%	5.45%	21.78%	63.62%	30.93%
2011	17,886.66	4,470.97	405	1,042.10	12,031.24	9.76%	5.83%	23.31%	61.44%	32.74%
2012	20,352.31	4,965.17	390	1,265.24	13,535.10	8.42%	6.22%	25.48%	60.29%	33.50%
2013	26,285.95	6,776.20	377	1,864.70	18,110.28	9.45%	7.09%	27.52%	61.80%	31.10%

Table 1.2: Descriptive statistics of price efficiency measures and control variables

The quarterly sample includes S&P 500 constituents and non-S&P 500 stocks that meet the minimum requirements for size and turnover ratio. *PO* is the fraction of share outstanding of a stock owned by passive funds. *NPO* is the fraction of share outstanding of a stock owned by non-passive institutional investors. *PT* is the sum of absolute passive holding changes of a stock during a quarter scaled by the total shares outstanding. *NPT* is the sum of absolute non-passive institutional holding changes of a stock during a quarter scaled by the total shares outstanding. *V*(s) is the pricing error of Hasbrouck (1993) estimated over a quarter and *V*(s)/*V*(p) is the relative pricing error (scaled by standard deviation of log price, *V*(p), over that quarter). *AC*(1) is the first-order autocorrelation of daily stock return. *VR*(1,5) is the ratio of weekly stock return variance to five times daily stock return variance. $\phi(OIB)$ is the coefficient of regressing 5-minute quote-midpoint return on lagged 5-minute volume-based order imbalance. *CAR*(-2, 1) is the cumulative abnormal return for a 4-day window around an earnings announcement. *CAR*(2, 60) is the cumulative abnormal return from the 2nd to the 60th day after an earnings announcement. *SUE* is the standardized unexpected earnings. *AFE* is the standardized analyst earnings forecast error. *RES* is size-weighted relative effective spread and *RQS* is time-weighted relative quote spread. *Amihud* is the Amihud (2002) price impact measure of illiquidity. *Liu* is the Liu (2006) no-trade-day measure. *MV* is the market value of stock estimated by the approach of Fama and French (1992). *TO* is the turnover ratio of the previous quarter. *PRC* is the share price at quarter end. *ANLY* is number of analyst following.

	2002-2013		2002-2007		2008-2013	
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
Average number of stocks per quarter	594		598		590	
Measures of Efficiency						
$V(s)/V(p)$ ($\times 10^{-2}$)	0.007	0.009	0.008	0.011	0.004	0.007
$\|AC(1)\|$	0.106	0.081	0.107	0.081	0.106	0.082
$\|1-VR(1,5)\|$	0.286	0.200	0.283	0.202	0.288	0.199
$\|\phi(OIB)\|$ ($\times 10^{-3}$)	0.153	0.166	0.129	0.138	0.177	0.188
Ownership & Trading						
PO	5.65%	3.14%	4.25%	1.91%	7.06%	3.48%
NPO	69.56%	17.98%	69.27%	17.75%	69.85%	18.21%
Quarterly PT	0.94%	1.23%	0.70%	0.88%	1.19%	1.46%
Quarterly NPT	20.69%	13.86%	20.65%	13.92%	20.72%	13.79%
Post-Earnings-Ann. Drift						
CAR(-2,1)	0.14%	6.18%	0.22%	5.82%	0.06%	6.52%
CAR(2,60)	0.36%	12.56%	0.34%	12.21%	0.39%	12.91%
SUE	-0.03%	2.40%	0.13%	1.93%	-0.18%	2.76%
AFE	0.07%	0.42%	0.06%	0.34%	0.08%	0.49%
Other Variables						
Illiquidity						
RES	0.17%	0.13%	0.20%	0.15%	0.13%	0.10%
RQS	0.66%	0.46%	0.71%	0.45%	0.61%	0.46%
Amihud	0.029	0.036	0.034	0.041	0.023	0.030
Liu	0.013	0.007	0.016	0.008	0.010	0.005
MV ($ billion)	15.39	21.18	14.60	20.50	16.18	21.81
PRC ($)	47.57	48.29	44.23	35.84	50.91	57.98
TO (/quarter)	72.69%	55.20%	58.72%	47.14%	86.71%	59.01%
ANLY	13.81	8.11	12.59	7.44	15.03	8.56

Table 1.3: Impact of passive ownership on post-earnings-announcement drift

Observations of quarterly earnings announcements are pooled over the period from 2002 to 2013 based on a stock sample including S&P 500 constituents and non-S&P 500 stocks that meet the minimum requirements for size and turnover ratio. The quarterly decile rank of the cumulative abnormal returns (*CAR*) from the 2^{nd} to the 10^{th}, 30^{th}, and 60^{th} day after an earnings announcement are regressed on earnings surprise and its interaction term with other variables. Abnormal returns are defined as size- and book-to-market-adjusted returns. The quarterly decile rank of cumulative abnormal return for a 4-day window around an earnings announcement, *DCAR*(-2,1), is used as proxy for earnings surprise. *DPO*, *DNPO*, *DMV*, *DPRC*, *DRES*, and *DVOL* are the decile portfolios of *PO*, *NPO*, *MV*, *PRC*, *RES* and *VOL* of the prior quarter, respectively. All of the decile portfolios are scaled to range between zero and one.

Dependent variable	*DCAR(2,10)*		*DCAR(2,30)*		*DCAR(2,60)*	
	[1]	[2]	[3]	[4]	[5]	[6]
DCAR(-2,1)	0.034	0.048	0.047	0.046	0.024	0.016
t-statistic	5.49	2.86	7.50	2.76	3.87	0.98
*DCAR(-2,1)*DPO*		0.035		0.017		0.022
t-statistic		3.52		1.73		2.19
*DCAR(-2,1)*DNPO*		-0.028		-0.013		-0.025
t-statistic		-2.52		-1.18		-2.26
*DCAR(-2,1)*DMV*		-0.007		0.006		0.010
t-statistic		-0.44		0.36		0.67
*DCAR(-2,1)*DPRC*		0.001		-0.008		-0.008
t-statistic		0.08		-0.63		-0.57
*DCAR(-2,1)*DRES*		-0.017		0.026		0.045
t-statistic		-1.67		2.50		4.35
*DCAR(-2,1)*DVOL*		-0.008		-0.028		-0.032
t-statistic		-0.48		-1.75		-1.98
N	25,269	24,610	25,228	24,570	25,036	24,382
Adj. R^2	0.001	0.002	0.002	0.003	0.001	0.002

Table 1.4: Robustness tests on post-earnings-announcement drift

Observations of quarterly earnings announcements are pooled over the period from 2002 to 2013. *DCAR*, *DRET*, *DPO*, *DNPO*, *DMV*, *DPRC*, *DRES*, *DVOL*, *DAFE*, and *DSUE* are the decile portfolios of *CAR*, *RET*, *PO*, *NPO*, *MV*, *PRC*, *RES*, *VOL*, *AFE*, and *SUE* of the prior quarter, respectively. All of the decile portfolios are scaled to range between zero and one. In Panel A, analyst forecast error (*AFE*) is used as proxy for earnings surprise. In Panel B, standardized unexpected earnings (*SUE*), defined as seasonal difference in announced earnings standardized by stock price at the end of the previous fiscal quarter, is used as proxy for earnings surprise. In Panel C, abnormal return is defined as size-adjusted return. In Panel D, raw returns are used to replace abnormal returns.

Dependent variable	*DCAR(2,10)*		*DCAR(2,30)*		*DCAR(2,60)*	
	[1]	[2]	[3]	[4]	[5]	[6]
Panel A: Earnings surprise is measured by analyst forecast error (*AFE*).						
DAFE	0.038	0.053	0.022	0.076	0.010	0.037
t-statistic	5.99	2.07	3.46	2.98	1.52	1.45
*DAFE*DPO*		0.033		0.015		0.028
t-statistic		3.16		1.45		2.66
*DAFE*DNPO*		-0.036		-0.018		-0.030
t-statistic		-3.20		-1.63		-2.68
Panel B: Earnings surprise is measured by standardized unexpected earnings (*SUE*).						
DSUE	0.024	0.031	0.020	0.058	0.011	0.035
t-statistic	3.76	1.24	3.20	2.32	1.75	1.41
*DSUE*DPO*		0.025		0.013		0.026
t-statistic		2.42		1.27		2.55
*DSUE*DNPO*		-0.017		-0.004		-0.019
t-statistic		-1.58		-0.38		-1.76
Panel C: *CAR* is based on size-adjusted returns.						
DCAR(-2,1)	0.034	0.040	0.048	0.057	0.032	0.028
t-statistic	5.59	1.64	7.91	2.31	5.20	1.14
*DCAR(-2,1)*DPO*		0.029		0.015		0.019
t-statistic		2.98		1.48		1.96
*DCAR(-2,1)*DNPO*		-0.023		-0.017		-0.027
t-statistic		-2.16		-1.61		-2.56
Panel D: Using raw return instead of *CAR* for both dependent and independent variables.						
DRET(-2,1)	0.089	0.207	0.082	0.278	0.033	0.040
t-statistic	13.53	7.66	12.29	10.04	4.74	1.39
*DRET(-2,1)*DPO*		0.022		0.003		0.049
t-statistic		2.08		0.27		4.37
*DRET(-2,1)*DNPO*		-0.013		-0.011		-0.033
t-statistic		-1.17		-0.94		-2.68

Table 1.5: Cross-sectional relation between passive ownership and price efficiency

The quarterly sample includes S&P 500 constituents and non-S&P 500 stocks that meet the minimum requirements for size and turnover ratio from 2002 to 2013. Measure of price efficiency is regressed on lagged passive ownership (*PO*) and control variables. Following the Fama and MacBeth (1973) approach, cross-sectional regressions are estimated in each quarter and the mean coefficients are reported. *V*(s) is the pricing error of Hasbrouck (1993) estimated over a quarter and *V*(s)/*V*(p) is the relative pricing error (scaled by standard deviation of log price, *V*(p), over that quarter). *AC*(1) is the first-order autocorrelation of daily stock return. *VR*(1,5) is the ratio of weekly stock return variance to five times daily stock return variance. $\phi(OIB)$ is the coefficient of regressing 5-minute quote-midpoint return on lagged 5-minute volume-based order imbalance. PO_{t-1} is the percentage of shares outstanding held by our passive fund sample at the end of the previous quarter. NPO_{t-1} is the percentage of shares outstanding held by non-passive institutional investors at the end of the previous quarter. RES_{t-1} is the trade-weighted relative effective spread estimated from the previous quarter. TO_{t-1} is the turnover ratio over the previous quarter. *MV* is the market value estimated by the approach of Fama and French (1992). PRC_{t-1} is the stock price at the end of the previous quarter. All variables are standardized to have zero-mean and unit standard deviation in each quarter. The significance level is based on the time-series variation in the quarterly regression coefficients over the sample period. The standard errors are adjusted for residual autocorrelation and heteroskedasticity based on Newey and West (1987).

Dependent variable	Ln[*V*(s)/*V*(p)]	|*AC*(1)|	|1-*VR*(1,5)|	Ln[|$\phi(OIB)$|]
Intercept	0.00	0.00	0.00	0.00
t-statistic	-0.53	-2.06	-0.53	-3.04
PO_{t-1}	0.06	0.04	0.03	0.03
t-statistic	3.36	3.39	2.36	3.10
NPO_{t-1}	-0.10	-0.01	-0.02	-0.02
t-statistic	-7.88	-1.26	-1.70	-2.46
Ln[RES_{t-1}]	0.07	0.00	0.01	0.19
t-statistic	1.73	-0.46	0.79	11.74
Ln[*MV*]	0.02	-0.02	-0.01	0.10
t-statistic	1.30	-2.47	-1.27	8.50
Ln[PRC_{t-1}]	-0.50	0.00	-0.01	-0.01
t-statistic	-21.13	-0.05	-0.81	-1.01
Ln[TO_{t-1}]	-0.16	-0.05	-0.02	0.12
t-statistic	-7.46	-3.74	-2.98	9.65
Average *N*	600	602	602	581
Adj. R^2	0.33	0.02	0.02	0.07

Table 1.6: Endogeneity

The quarterly sample includes S&P 500 constituents and non-S&P 500 stocks that meet the minimum requirements for size and turnover ratio from 2002 to 2013. *V*(s) is the pricing error of Hasbrouck (1993) estimated over a quarter and *V*(s)/*V*(p) is the relative pricing error (scaled by standard deviation of log price, *V*(p), over that quarter). *AC*(1) is the first-order autocorrelation of daily stock return. *VR*(1,5) is the ratio of weekly stock return variance to five times daily stock return variance. $\phi(OIB)$ is the coefficient of regressing 5-minute quote-midpoint return on lagged 5-minute volume-based order imbalance. PO_{t-1} is the percentage of shares outstanding held by our passive fund sample at the end of the previous quarter, and ΔPO_{t-1} is its change from quarter (*t*-2) to (*t*-1). NPO_{t-1} is the percentage of shares outstanding held by non-passive institutional investors at the end of the previous quarter, and ΔNPO_{t-1} is its change from quarter (*t*-2) to (*t*-1). $Efficiency_{t-1}$ is the lagged price inefficiency measure. MV_{t-1} is the stock's market value at the end of the previous quarter. *SP_ADD* equals to 1 if the stock is added to the S&P 500 index at that quarter and zero otherwise. *SP_DEL* equals to 1 if the stock is deleted from the S&P 500 index at that quarter and zero otherwise. Panel A repeats the cross-sectional regression in Table 3 but includes only pure index funds and ETFs into the passive fund sample. Coefficients of control variables are omitted for brevity. Panel B presents the results of cross-sectional regression of change in *PO* on lagged change in price efficiency measures and other control variables. The significance level is based on the time-series variation in the quarterly regression coefficients over the sample period. The standard errors are adjusted for residual autocorrelation and heteroskedasticity based on Newey and West (1987).

Panel A: Passive fund sample includes only pure index funds and ETFs.

Dependent variable	Ln[*V*(s)/*V*(p)]	|*AC*(1)|	|1-*VR*(1,5)|	Ln[|ϕ(*OIB*)|]
PO_{t-1}	0.08	0.04	0.03	0.03
t-statistic	5.02	2.98	2.27	2.87
NPO_{t-1}	-0.11	-0.01	-0.02	-0.01
t-statistic	-7.05	-1.10	-1.58	-2.26

Panel B: Cross-sectional determinants of change in *PO*.

Measure of Price Inefficiency	ΔLn[*V*(s)/*V*(p)]	Δ|*AC*(1)|	Δ|1-*VR*(1,5)|	ΔLn[|ϕ(*OIB*)|]
Intercept	-0.02	-0.02	-0.02	-0.01
t-statistic	-4.28	-4.23	-4.25	-1.56
$\Delta Efficiency_{t-1}$	-0.02	0.00	0.01	0.00
t-statistic	-1.71	-0.21	1.41	-0.28
ΔPO_{t-1}	-0.10	-0.10	-0.10	-0.10
t-statistic	-5.29	-5.23	-5.25	-5.50
$\Delta \text{Ln}[RES_{t-1}]$	0.00	0.00	0.00	0.00
t-statistic	0.17	0.33	0.51	-0.52
$\Delta \text{Ln}[MV_{t-1}]$	0.02	0.02	0.02	0.02
t-statistic	1.73	1.69	1.80	1.58
$\Delta \text{Ln}[TO_{t-1}]$	0.00	0.00	0.00	0.00
t-statistic	-0.09	0.51	0.46	0.23
SP_ADD	1.46	1.48	1.48	1.45
t-statistic	7.84	8.07	8.07	6.99
SP_DEL	-1.15	-1.15	-1.15	-1.11
t-statistic	-3.98	-4.04	-3.98	-3.18
N	602	602	602	602
Adj R^2	0.12	0.12	0.12	0.13

Table 1.7: Robustness tests on cross-section analysis

The quarterly sample includes S&P 500 constituents and non-S&P 500 stocks that meet the minimum requirements for size and turnover ratio from 2002 to 2013. The same Fama-MacBeth regressions as in Table 3 are estimated, but coefficients of control variables are omitted for brevity. *V*(s) is the pricing error of Hasbrouck (1993) estimated over a quarter and *V*(s)/*V*(p) is the relative pricing error (scaled by standard deviation of log price, *V*(p), over that quarter). *AC*(1) is the first-order autocorrelation of daily stock return. *VR*(1,5) is the ratio of weekly stock return variance to five times daily stock return variance. $\phi(OIB)$ is the coefficient of regressing 5-minute quote-midpoint return on lagged 5-minute volume-based order imbalance. *RQS* is the time-weighted relative quoted spread. *Amihud* is the Amihud (2002) price impact measure of illiquidity. *Liu* is the Liu (2006) no-trade-day measure. Panel A presents results by different liquidity measures. Panel B presents results by dropping the liquidity control variable. Panel C tests several different specifications. All variables are standardized to have zero-mean and unit standard deviation in each quarter. The significance level is based on the time-series variation in the quarterly regression coefficients over the sample period. The standard errors are adjusted for residual autocorrelation and heteroskedasticity based on Newey and West (1987).

Dependent variable	Ln[*V*(s)/*V*(p)]	|*AC*(1)|	|1-*VR*(1,5)|	Ln[|ϕ(*OIB*)|]
Panel A: Alternative liquidity measures				
PO_{t-1}	0.05	0.04	0.03	0.00
t-statistic	2.19	2.98	2.19	-0.29
NPO_{t-1}	-0.09	-0.01	-0.02	-0.03
t-statistic	-8.10	-1.07	-1.50	-3.38
Ln[RQS_{t-1}]	-0.05	-0.03	-0.01	0.16
t-statistic	-1.37	-2.55	-0.92	11.62
PO_{t-1}	0.03	0.02	0.12	0.00
t-statistic	1.92	2.54	2.04	2.14
NPO_{t-1}	-0.09	-0.01	-0.02	-0.01
t-statistic	-5.23	-1.22	-1.47	-1.50
Ln[$Amihud_{t-1}$]	-0.47	-0.12	-0.02	0.28
t-statistic	-10.17	-4.28	-1.27	11.75
PO_{t-1}	0.06	0.04	0.03	0.00
t-statistic	3.70	3.47	2.22	0.08
NPO_{t-1}	-0.10	-0.01	-0.02	-0.01
t-statistic	-7.15	-1.29	-1.55	-1.08
Ln[Liu_{t-1}]	0.26	0.08	0.09	-0.15
t-statistic	4.09	2.19	1.51	-3.07
Panel B: Drop liquidity control variables				
PO_{t-1}	0.06	0.04	0.03	0.00
t-statistic	3.55	3.38	2.12	0.13
NPO_{t-1}	-0.10	-0.01	-0.02	-0.01
t-statistic	-7.25	-1.31	-1.54	-1.01

Table 1.7 (continued)

Dependent variable	Ln[V(s)/V(p)]	|AC(1)|	|1-VR(1,5)|	Ln[|ϕ(OIB)|]
Panel C: Other alternative specifications				
Add S&P 500 Indicator				
PO_{t-1}	0.09	0.03	0.03	0.02
t-statistic	4.24	2.48	2.03	2.33
NPO_{t-1}	-0.10	-0.01	-0.02	-0.02
t-statistic	-7.91	-1.28	-1.69	-2.26
S&P 500 Dummy	-0.09	0.00	-0.01	0.03
t-statistic	-0.99	0.12	-0.67	1.13
Controlling for lagged dependent variable				
PO_{t-1}	0.03	0.03	0.03	0.03
t-statistic	3.44	3.54	2.40	3.34
NPO_{t-1}	-0.05	-0.01	-0.02	-0.02
t-statistic	-6.88	-1.17	-1.71	-2.27
DV_{t-1}	0.50	0.03	0.01	0.06
t-statistic	16.79	2.96	2.54	11.24
Use contemporaneous variables				
PO_t	0.06	0.04	0.03	0.03
t-statistic	4.27	3.36	1.97	4.03
NPO_t	-0.10	-0.01	-0.02	-0.01
t-statistic	-7.86	-1.12	-0.81	-2.66
Drop stock-quarters with beginning price lower than $5				
PO_{t-1}	0.06	0.04	0.03	0.03
t-statistic	3.25	3.43	2.45	3.05
NPO_{t-1}	-0.10	-0.01	-0.02	-0.01
t-statistic	-7.74	-1.39	-1.67	-1.90

Table 1.8: Relation between passive ownership and analyst following

The quarterly sample includes all S&P 500 constituents from 2002 to 2013 except for the ones at the bottom size decile. Natural logarithm of number of analyst following is regressed on lagged passive ownership and control variables. Following the Fama and MacBeth (1973) approach, cross-sectional regressions are estimated in each quarter and the mean coefficients are reported. PO_{t-1} is the percentage of shares outstanding held by our passive fund sample at the end of the previous quarter. NPO_{t-1} is the percentage of shares outstanding held by non-passive institutional investors at the end of the previous quarter. $ANLY_{t-1}$ is the lagged dependent variable. *TO* is the turnover ratio over the previous year. *MV* is the market value of the stock estimated by the approach of Fama and French (1992). $SP500D_{t-1}$ is an indicator variable that equals to 1 if the stock is a member of the S&P 500 index in that quarter and zero otherwise. $R1000D_{t-1}$ is an indicator variable that equals to 1 if the stock is a member of the Russell 1000 index in that quarter and zero otherwise. All variables are standardized to have zero-mean and unit standard deviation in each quarter. The significance level is based on the time-series variation in the quarterly regression coefficients over the sample period. The standard errors are adjusted for residual autocorrelation and heteroskedasticity based on Newey and West (1987).

	[1]	[2]	[3]
Intercept	-1.58	-1.53	-0.15
t-statistic	-11.66	-14.77	-10.11
PO_{t-1}	-0.12	-0.12	-0.01
t-statistic	-2.68	-6.13	-5.89
NPO_{t-1}	0.07	0.04	0.00
t-statistic	7.99	3.52	2.28
Ln[$ANLY_{t-1}$]			0.91
t-statistic			64.62
Ln[*TO*]		0.24	0.03
t-statistic		10.71	4.52
Ln[*MV*]		0.25	0.03
t-statistic		16.60	4.80
$SP500D_{t-1}$	0.46	0.26	0.02
t-statistic	17.11	7.57	2.64
$R1000D_{t-1}$	1.45	1.56	0.16
t-statistic	9.36	11.47	10.68
Average *N*	602	602	602
Adj. R^2	0.33	0.42	0.89

Table 1.9: Effect of passive ownership and trading on price efficiency

The quarterly sample includes matched S&P 500 constituents and non-S&P 500 stocks from 2002 to 2013. The dependent variable is one of the five price inefficiency measures. *V*(s) is the pricing error of Hasbrouck (1993) estimated over a quarter and *V*(s)/*V*(p) is the relative pricing error (scaled by standard deviation of log price, *V*(p), over that quarter). *AC*(1) is the first-order autocorrelation of daily stock return. *VR*(1,5) is the ratio of weekly stock return variance to five times daily stock return variance. $\phi(OIB)$ is the coefficient of regressing 5-minute quote-midpoint return on lagged 5-minute volume-based order imbalance. PT_t is the sum of absolute passive holding changes of a stock during the current quarter scaled by the total shares outstanding. NPT_t is the sum of absolute non-passive institutional holding changes of a stock during the current quarter scaled by the total shares outstanding. PO_{t-1} is the percentage of shares outstanding held by the passive fund sample at the end of the previous quarter. NPO_{t-1} is the percentage of shares outstanding held by non-passive institutional investors at the end of the previous quarter. RES_{t-1} is the trade-weighted relative effective spread of previous quarter. TO_{t-1} is the turnover ratio over the previous quarter. *MV* is the market value estimated by the approach of Fama and French (1992). PRC_{t-1} is stock price at the end of the previous quarter. All variables are standardized to have zero-mean and unit standard deviation in each quarter. Following the Fama and MacBeth (1973) approach, cross-sectional regressions are conducted in each quarter and the mean coefficients are reported. The significance level is based on the time-series variation in the quarterly regression coefficients over the sample period. The standard errors are adjusted for residual autocorrelation and heteroskedasticity based on Newey and West (1987).

Dependent variable	Ln[*V*(s)/*V*(p)]	|*AC*(1)|	|1-*VR*(1,5)|	Ln[|$\phi(OIB)$|]
Intercept	0.06	-0.01	-0.03	0.02
t-statistic	4.27	-0.64	-1.31	0.85
PT_t	0.02	0.00	0.00	0.01
t-statistic	1.70	0.45	0.50	0.93
NPT_t	-0.28	0.05	0.13	-0.12
t-statistic	-4.00	0.71	1.37	-0.95
PO_{t-1}	0.05	0.04	0.03	0.03
t-statistic	2.36	3.41	2.66	3.52
NPO_{t-1}	-0.09	-0.01	-0.02	-0.02
t-statistic	-7.79	-1.19	-2.07	-1.73
Ln[RES_{t-1}]	0.08	0.00	0.01	0.18
t-statistic	1.84	-0.48	0.96	11.14
Ln[*MV*]	0.02	-0.02	-0.01	0.10
t-statistic	1.16	-2.39	-0.91	7.89
Ln[PRC_{t-1}]	-0.50	0.00	-0.01	-0.02
t-statistic	-21.41	-0.01	-0.96	-1.08
Ln[TO_{t-1}]	-0.16	-0.05	-0.03	0.12
t-statistic	-8.18	-3.69	-3.46	9.47
Average *N*	600	602	602	581
Adj. R^2	0.33	0.03	0.02	0.07

Table 1.10: Effects of passive ownership and trading on liquidity

The quarterly sample includes matched S&P 500 constituents and non-S&P 500 stocks from 2002 to 2013. Dependent variable is one of the four illiquidity measures. *RES* is trade-weighted relative effective spread and *RQS* is time-weighted relative quoted spread. *Amihud* is the Amihud (2002) price impact measure. *Liu* is the Liu (2006) no-trade-day measure. *TO* is the quarterly turnover ratio. PO_{t-1} is the percentage of shares outstanding held by the passive fund sample at the end of the previous quarter. NPO_{t-1} is the percentage of shares outstanding held by non-passive institutional investors at the end of the previous quarter. PT_t is the sum of absolute passive holding changes of a stock during the current quarter scaled by the total shares outstanding. NPT_t is the sum of absolute non-passive institutional holding changes of a stock during the current quarter scaled by the total shares outstanding. *MV* is the market value estimated by the approach of Fama and French (1992). PRC_{t-1} is stock price at the end of the previous quarter. $SP500D_{t-1}$ is equal to 1 if the stock is currently a member of the S&P 500 index and 0 otherwise. All variables are standardized to have zero-mean and unit standard deviation in each quarter. The significance level is based on the time-series variation in the quarterly regression coefficients over the sample period. The standard errors are adjusted for residual autocorrelation and heteroskedasticity based on Newey and West (1987).

Dependent variable	Ln[*RES*]	Ln[*RQS*]	Ln[*Amihud*]	Ln[*Liu*]	Ln[*TO*]
Panel A: Effect of Ownership on liquidity					
Intercept	0.04	-0.22	0.04	-0.08	0.09
t-statistic	0.67	-3.04	1.16	-1.33	1.46
PO_{t-1}	-0.10	-0.03	-0.02	-0.01	0.01
t-statistic	-2.06	-0.30	-0.74	-0.21	0.20
NPO_{t-1}	0.06	0.06	-0.04	-0.18	0.17
t-statistic	4.29	3.84	-3.11	-15.00	14.99
Ln[*MV*]	-0.27	-0.40	-0.80	0.18	-0.18
t-statistic	-4.84	-9.41	-58.94	13.58	-12.78
Ln[PRC_{t-1}]	-0.03	0.08	-0.12	0.08	-0.09
t-statistic	-0.30	1.56	-4.38	3.34	-3.97
$SP500D_{t-1}$	-0.06	0.33	-0.07	0.13	-0.14
t-statistic	-0.73	2.96	-1.26	1.33	-1.48
Average *N*	601	590	602	602	602
Adj. R^2	0.27	0.27	0.75	0.13	0.13

Table 1.10: Effects of passive ownership and trading on liquidity (continued)

Dependent variable	Ln[*RES*]	Ln[*RQS*]	Ln[*Amihud*]	Ln[*Liu*]	Ln[*TO*]
Panel B: Effect of ownership and trading on liquidity					
Intercept	0.04	-0.20	0.06	-0.04	0.05
t-statistic	0.73	-3.05	1.79	-0.70	0.84
PT_t	0.00	-0.03	-0.02	-0.03	0.02
t-statistic	-0.03	-1.33	-1.10	-1.12	0.65
NPT_t	0.05	-0.09	-0.18	-0.52	0.54
t-statistic	1.86	-4.08	-11.00	-38.33	41.11
PO_{t-1}	-0.09	-0.02	-0.03	-0.05	0.06
t-statistic	-2.06	-0.28	-1.25	-0.85	0.92
NPO_{t-1}	0.04	0.09	0.03	0.01	-0.02
t-statistic	2.50	5.53	2.12	0.74	-1.28
Ln[*MV*]	-0.26	-0.42	-0.85	0.05	-0.05
t-statistic	-4.55	-10.09	-48.76	3.73	-3.57
Ln[PRC_{t-1}]	-0.03	0.08	-0.13	0.06	-0.07
t-statistic	-0.29	1.52	-4.40	2.82	-3.32
$SP500D_{t-1}$	-0.06	0.30	-0.09	0.05	-0.07
t-statistic	-0.78	2.95	-1.93	0.63	-0.79
Average *N*	601	590	602	602	602
Adj. R^2	0.29	0.30	0.79	0.35	0.37

Essay 2: Effect of Price Inefficiency on Idiosyncratic Risk and Stock Returns

1. Introduction

Classical asset pricing theories rely on portfolio diversification and imply informational efficiency of stock prices. As random mispricing of individual stocks does not exist in equilibrium and, even if exists in very short-horizon, will not affect performance of a well-diversified portfolio, price efficiency is not modeled in pricing theories for investor decisions or cross-sectional variation in expected returns. Increasing evidence, however, indicates the presence of price inefficiency and its profound impact on financial markets[22], as well as potential underdiversification among investors.[23] In scenarios of underdiversification, asset valuation will be affected by price inefficiency.

This essay integrates the market efficiency and the asset pricing literatures, presenting evidence that price inefficiency could have significant effect on the cross-section of stock prices. It is based on the notion that price inefficiency adds idiosyncratic uncertainty to stock returns. Given the lack of diversification, price inefficiency through its effect on idiosyncratic risk could be related to stock returns if investors demand extra returns for holding idiosyncratic risk (Levy, 1978; Merton, 1987).

This essay first proposes a simple model to quantify the impact of price inefficiency on expected idiosyncratic volatility. In this model, price is an unbiased estimator of fundamental value but random pricing error could exist. The pricing error is assumed to be a log-normally distributed random variable with zero mean and positive, time-varying volatility. It has zero cross-sectional and serial correlations and is independent of systematic risk. The model indicates that random pricing error adds extra idiosyncratic

[22] For example, De Bondt and Thaler (1985) find evidence of stock return momentum and reversals; Black (1986) notes that mispricing probably plays a profound role in financial markets; French and Roll (1986) find that approximately 4% to 12% of the daily stock return variance comes from mispricing; Barberis, Shleifer, and Vishny (1998) propose a model of investor sentiment to explain empirical findings of underreaction and overreaction.

[23] Barber and Odean (2000) report an average of only four stocks in the portfolio of a typical U.S. individual investor; Polkovnichenko (2005) and Goetzmann and Kumar (2008) also provide similar evidence.

volatility (IVOL) to future stock returns. The magnitude of this extra IVOL depends mainly on the magnitude of the pricing error.

Using deviations of stock price from random walk as proxies for price inefficiency, this essay then provides empirical evidence that both absolute and relative IVOL in U.S. common stocks increase with lagged empirical measures of price inefficiency, consistent with the prediction of the model. Next, this essay demonstrates a positive relation between expected returns and price inefficiency in U.S. stocks by both cross-sectional regressions and zero-investment portfolios. This return premium of price inefficiency is both statistically and economically significant: a value-weighted zero-investment portfolio that is long in the top quintile and short in the bottom quintile of stocks sorted by price inefficiency earns a large annualized four-factor-adjusted return of 6% over the period from 1984 to 2012. Moreover, this price inefficiency premium disappears, or largely declines, in stocks with high institutional ownership. This is consistent with the hypothesis that idiosyncratic risk is priced when and only when shareholders are underdiversified.

Moreover, using effective spreads or quoted spreads as measures of transactions costs, this essay finds that trading strategies based on price inefficiency could realize significant abnormal returns even after deducting transactions costs. Finally, this essay finds that the price inefficiency premium can entirely explain the size effect, though it has little explanatory power for the value and momentum effects.

To the best of my knowledge, this is the first study where price inefficiency affects cross-sectional stock returns due to its impact on expected idiosyncratic risk. It is evidently different from the bid-ask or mispricing premium proposed by Blume and Stambaugh (1983) and Brennan and Wang (2010). Blume and Stambaugh (1983) argue that bid-ask bounce generates an upward bias on stock returns computed with closing prices, due to the non-linear relation between price and expected return. Brennan and Wang (2010) extend and generalize this Jensen's inequality argument, theoretically as well as empirically, by pointing out that pricing errors can generate an upward bias in unconditional expected returns. The underpricing in some stocks and overpricing in other stocks causes equally-weighted returns to be upwardly biased due to Jensen's inequality. However, returns of value-weighted (VW) portfolios used in this essay are largely unaffected by Jensen's inequality, though they could be affected by extra idiosyncratic risk caused by price inefficiency as far as idiosyncratic risk is priced. In other words, while Jensen's inequality mechanically produces upward biases in expected *rates of returns* of mispriced stocks, the results in this essay represent higher expected *values* of low-efficiency stocks due to greater idiosyncratic risk.

This essay is related to De Long et al. (1990) who propose that noise trader risk could generate higher expected returns. The price inefficiency examined by this essay should be typically caused by noise traders. Moreover, both papers imply a positive relation between uninformed trading and expected returns. However, this essay has a different mechanism that generates this positive relation. In De Long et al. (1990), short-horizon arbitrageurs are unwilling to hold mispriced assets due to noise trader risk–misperceptions of noise traders could become even more extreme in future than they are at present. In this essay, however, expected IVOL is increased not only when noise traders generate price inefficiency but also when the market corrects the pricing errors. Underdiversified investors thus are unwilling to hold low-price-efficiency assets even if they expect quick disappearance of the pricing errors. In other words, inferences of this essay do not rely on the noise trader risk of De Long et al. (1990), though the inferences are complementary.

This essay naturally contributes to the debate on whether idiosyncratic risk is priced. Merton (1987) proposes that idiosyncratic risk should be priced due to underdiversification of investors' stock holdings. However, Ang et al. (2006), among several others, reveal a puzzling phenomenon that stocks with high IVOL have lower average returns in future. Several recent studies have proposed explanations for this puzzle. For example, Fu (2009), Huang et al. (2010), and Chua, Goh, and Zhang (2010) argue that findings of Ang et al. (2006) are largely explained by the return reversals following high stock returns produced by (unexpected) contemporaneous IVOL shock. Chen and Petkova (2012) suggest that portfolios with high IVOL have exposure to innovations in average stock variance and thus lower expected returns. The common ground of these explanations is that lagged IVOL is not a proper measure of expected IVOL and could bring complications such as return reversals or missing-factor problem. The appealing feature of this essay is that it provides a setting to test the relation between *expected* IVOL and stock returns without introducing too much cross-sectional variation in *lagged* IVOL, thus could largely alleviate the complications mentioned above. The return premium associated with price inefficiency supports the notion that idiosyncratic risk is priced, and the disappearance of this premium in stocks with high institutional ownership is also consistent with Merton (1987).

The rest of this essay is organized as follows. Section 2 presents a simple model of the effect of price inefficiency on expected idiosyncratic volatility. Section 3 introduces the data, proxies for price inefficiency, and other related measures. Section 4 empirically investigates the relations between price inefficiency, future IVOL and future stock returns. Section 5 conducts extensive robustness tests. Section 6 explores whether the price inefficiency premium could be realized net of transaction costs and whether it contributes to traditional risk factors. Section 7 concludes.

2. A simple model of price-inefficiency-induced idiosyncratic risk

The model of price-inefficiency-induced volatility is a simplified variant of a multi-period framework used in Brennan and Wang (2010) and several other studies the literature of mispricing. The market price of a stock at the end of period t is denoted by P_t. The unobservable efficient price, which incorporates all available public and private information, is denoted by P_t^*. Define $Z_t = P_t/P_t^*$ as a measure of the (relative) pricing error. It is assumed to have zero cross-sectional and serial correlations[24] and is independent from changes in fundamental price. The market price is assumed to be an unbiased estimate of efficient price so that $E[Z_t] = 1$. The natural logarithm of the pricing error, $z_t = \ln Z_t$, is assumed to be an independently and normally distributed random variable with time-varying variance, $z_t \sim N(\mu, \sigma_{z,t}^2)$, $\mu = -\sigma_{z,t}^2/2$. The variance, $\sigma_{z,t}^2$, serves as the measure

[24] A more realistic version of this model allows pricing error to have non-zero serial correlations. However, as mispricing is typically a temporary phenomenon, it is not likely to have significant serial correlation on the monthly frequency used in this essay. Moreover, as far as pricing errors are not perfectly and positively auto-correlated, pricing error will have positive impact on expected idiosyncratic volatility. The magnitude of this impact will be negatively associated with the serial correlation.

for the magnitude of the pricing error. For simplicity, the stock is assumed to pay no dividend.

Based on above settings, the gross holding period return of this stock over period *t* could be expressed as:

$$1 + r_t = \frac{P_t}{P_{t-1}} = \frac{P_t^* Z_t}{P_{t-1}^* Z_{t-1}} = \frac{(1 + r_t^*)Z_t}{Z_{t-1}} = (1 + r_t^*)e^{\Delta z_t}, \tag{2.1}$$

where $\Delta z_t = z_t - z_{t-1}$ is the change in log pricing error and follows a normal distribution, $\Delta z_{i,t} \sim N(0, \sigma_{z,t-1}^2 + \sigma_{z,t}^2)$, and $r_t^* = P_t^*/P_{t-1}^*$ is the fundamental return. As pricing error is usually a small fraction of stock price, expression of r_t could be simplified by power series approximation:

$$1 + r_t = (1 + r_t^*)e^{\Delta z_t} \approx (1 + r_t^*)(1 + \Delta z_t) = 1 + r_t^* + (1 + r_t^*)\Delta z_t. \tag{2.2}$$

Since pricing error is uncorrelated with efficient price, return variance could be expressed approximately by:

$$\begin{aligned} Var(r_t) &\approx Var(r_t^* + (1 + r_t^*)\Delta z_t) = Var(r_t^*) + Var\big((1 + r_t^*)\Delta z_t\big) \\ &= Var(r_t^*) + Var\big((1 + r_t^*)\Delta z_t\big) + 2Cov(r_t^*, r_t^*\Delta z_t) \\ &= Var(r_t^*) + \big((E[1 + r_t^*])^2 + Var(r_t^*)\big)Var(\Delta z_t) \\ &= Var(r_t^*) + 2\big((E[1 + r_t^*])^2 + Var(r_t^*)\big)\big(\sigma_{z,t-1}^2 + \sigma_{z,t}^2\big). \end{aligned} \tag{2.3}$$

where $Var(r_t^*)$ represents variance of the fundamental return. Therefore, the presence of price inefficiency increases stock return variance by approximately:

$$Var(r_t) - Var(r_t^*) \approx \big((E[1 + r_t^*])^2 + Var(r_t^*)\big)\big(\sigma_{z,t-1}^2 + \sigma_{z,t}^2\big). \tag{2.4}$$

At the beginning of period *t*, if investors use the magnitude of the realized pricing error ($\sigma_{z,t-1}^2$) to form the expectations of pricing error in period *t*, the price-inefficiency-induced expected variance will be:

$$E\{Var(r_t) - Var(r_t^*) \big| E\big[\sigma_{z,t}^2\big] = \sigma_{z,t-1}^2\} \approx 2((E[1 + r_t^*])^2 + E[Var(r_t^*)])\sigma_{z,t-1}^2. \tag{2.5}$$

Even if investors expect the pricing error to be completely corrected in period *t*, price inefficiency in *t*-1 still induces an extra expected variance of:

$$E\{Var(r_t) - Var(r_t^*)\big|E[\sigma_{z,t}^2] = 0\} \approx ((E[1 + r_t^*])^2 + E[Var(r_t^*)])\sigma_{z,t-1}^2. \quad (2.6)$$

As the fundament return, $1 + r_t^*$, and return induced by price inefficiency, $(1 + r_t^*)\Delta z_t$, have zero covariance, the price-inefficiency-induced volatility is completely idiosyncratic. Clearly, magnitude of this volatility depends on three factors. First and most importantly, there is an intuitive relation that greater magnitude of the realized pricing error, measured by $\sigma_{z,t-1}^2$, is associated with higher expected volatility. $\sigma_{z,t-1}^2$ is the single most important factor: if it becomes zero, there won't be any expected volatility in addition to that of the fundamental return. As $E[1 + r_t^*]$ is usually close to 1 and $E[Var(r_t^*)]$ generally has much smaller magnitude than $E[1 + r_t^*]$, the price-inefficiency-induced expected volatility should be close to $2\sigma_{z,t-1}^2$ when investors form their expectation of future price inefficiency based on the magnitude of the realized pricing errors. Second, the price-inefficiency-induced expected volatility is greater when the expected fundament return, $E[1 + r_t^*]$, becomes larger. The intuition is that as the expected stock price rises due to higher expected return, pricing error with the same proportion of the efficient price could cause greater magnitude of volatility. Finally, price-inefficiency-induced expected volatility also increases with the expected variance of the fundamental return, $E[Var(r_t^*)]$. However, the latter two factors are not as important as $\sigma_{z,t-1}^2$, even in the case when the fundamental return has zero mean and zero variance, price inefficiency still increases the return variance by approximately $2\sigma_{z,t-1}^2$.

3. Data and measures

3.1 The data

The empirical analysis of this essay is based on a sample of 9,004 U.S. common stocks that are listed on the NYSE and the AMEX from January 1984 to December 2012 and those listed on the NASDAQ from January 1993 to December 2012.[25] Penny stocks[26]

[25] NASDAQ stocks from 1984 to 1992 are not included due to data availability issue.
[26] As underpriced stocks tend to be small, removing stocks below certain size threshold could generate biased sample selection. To avoid this problem, I exclude stocks that are not within the Russell 3000 index, which

and stocks with no valid estimates of price inefficiency are excluded from the sample. Constituents of the S&P 500 index are obtained from Compustat and those of the Russell indices are obtained from Russell Investments. Monthly stock returns and characteristics are taken from the CRSP. Data on book value of equity are taken from Compustat. Intraday quote and trade data required in the estimation of inefficiency measures and transaction costs are taken from NYSE TAQ.[27] Data on analyst coverage are taken from the IBES. One-month T-bill rates and risk factors, including the three Fama-French (1993) factors, the momentum factor of Carhart (1997), and the illiquidity factor of Pastor-Stambaugh (2003), are obtained from WRDS at the University of Pennsylvania. Most variables used in this essay are winsorized by 0.5% at each tail.

Please insert Table 2.1 here

3.2 Measure of informational efficiency of stock price

As efficient price is, unfortunately, unobservable and difficult to estimate, it is impossible to precisely estimate σ_z^2 in the model proposed in Section 1. The use of empirical proxies of price inefficiency, which are usually noisy and controversial though, is a necessity. Proxies of price inefficiency fall into several groups. First, price inefficiency could be measured by the returns to trading strategies based on market anomalies, such as short-term reversals, momentums, and post-earnings announcement drift. Realization of abnormal returns could indicate the existence of price inefficiency. Second, under the assumption that efficient price follows a random walk, the deviations of stock prices from random walk could be a measure of relative inefficiency. This group of measures includes Hasbrouck (1993) pricing error variance, return autocorrelations and variance ratios. Third, price inefficiency could be measured by certain asset pricing models. For example, the Kalman filter estimation of mispricing used by Brennan and Wang (2010) is based on

accounts for over 98% of the U.S. investable equity market capitalization. Adding back these stocks to the sample does not change the results in this essay.

[27] Following Hasbrouck (1988) and Lee and Ready (1991), this essay assumes that trades are recorded five seconds late before 1998 and adjusts the time stamps accordingly. For the post-1998 period, the time stamps are assumed to be recorded correctly.

the assumption that fundament return follows an ex post version of the Fama-French (1993) three factor model. In this essay, I use Hasbrouck's (1993) pricing error variance as the main proxy for price inefficiency, and use weekly-to-daily variance ratio as alternative proxy in robustness tests.

3.1.1 Hasbrouck's (1993) pricing error

The pricing error proposed by Hasbrouck (1993) measures the deviation between transaction prices and an implicit random walk process, which is assumed to be the efficient price. Specifically, the log transaction price, p_t, is defined as the efficient price, m_t, plus a transitory deviation, s_t:

$$p_t = m_t + s_t. \tag{2.7}$$

t indexes transaction time; m_t is defined as the expectation of the stock value given all available public information and is assumed to follow a random walk; s_t measures the deviation of transaction price from the efficient price. Specifically, s_t is assumed to be a zero-mean covariance-stationary stochastic process with variance of σ_s^2. Clearly, σ_s^2 inversely measures how closely the transaction price follows the efficient price, thus is used as a measure of price inefficiency.

Following Hasbrouck (1993), I estimate the *lower bound* for σ_s^2 using a five-lag vector autoregression (VAR) model based on intraday trade and quote data obtained from NYSE TAQ.[28] However, estimation of σ_s^2 is largely affected by the return volatility, making the comparison of σ_s^2 across stocks much less meaningful.[29] To control for the cross-sectional difference in stock return variance, I follow Boehmer, Saar, and Yu (2005) and several other studies[30] to normalize σ_s^2 by the variance of the log transaction prices,

---

[28] See Appendix A for detailed estimation process.
[29] For example, in a simple version of Hasbrouck's (1993) variance decomposition that based only on returns (but not trade variables), the estimated σ_s^2 is a function of σ_p^2 and first-order autocorrelation in returns. Moreover, as the estimates of σ_s^2 from the variance decomposition is actually a lower bound of σ_s^2, it could be downward biased when σ_p^2 is low. In an extreme case where $\sigma_p^2 = 0$, the estimate of σ_s^2 will be zero regardless of whether pricing errors exist.
[30] Boehmer, Saar, and Yu (2005) use σ_s^2/σ_p^2 to study the effect of the increased pre-trade transparency on stock price efficiency. Boehmer and Kelley (2009) use this measure to study the effect of institutional ownership on stock price efficiency. Boehmer and Wu (2012) use this measure to examine the relation

σ_p^2, to form a relative measure of inefficiency, σ_s^2/σ_p^2. It represents the proportion of deviations from random walk in the total variability of the transaction return process and are used as the main proxy for inefficiency in this essay. Meanwhile, σ_s^2 is also used as an alternative proxy in robustness tests. As reported in Table 2.1, there are a total of 836,198 monthly estimates of σ_s^2/σ_p^2 with the mean of 1.29%, the median of 0.08%, and the skewness of 6.68.

3.1.2 Variance ratio

Variance ratio is a traditional measure of price inefficiency.[31] For a random walk process, ratio of long-interval and short-interval return variances pre unit time is equal to one. In the presence of pricing error, variance ratio could be above or below one, depending on the sign of return autocorrelation. Thus the deviation of variance ratio from one could be considered as a proxy of relative price inefficiency. Using the same approach of Lo and MacKinlay (1988), I estimate *VR*(q) for each stock-month based on daily quote-midpoint returns over the previous three months:[32]

$$\bar{\sigma}_a^2 = \frac{1}{nq-1}\sum_{k=1}^{nq}(X_k - X_{k-1} - \hat{\mu})^2, \tag{2.8}$$

$$\bar{\sigma}_c^2(q) = \frac{n}{(nq-q+1)(nq-q)}\sum_{k=q}^{nq}\left(X_k - X_{k-q} - q\hat{\mu}\right)^2, \tag{2.9}$$

$$VR(q) = \frac{\bar{\sigma}_c^2(q)}{\bar{\sigma}_a^2} - 1. \tag{2.10}$$

The absolute value of the weekly-to-daily variance ratio, |*VR*(5)|, is used as an alternative proxy for inefficiency in robustness test.[33]

---

between short selling and the price discovery process. Hotchkiss and Ronen (2002) use a variant of this measure, $MQ = 1 - 2\sigma_s^2/\sigma_p^2$, to examine the informational efficiency of corporate bond price.

[31] Another traditional measure, autocorrelation, is closely related to variance ratios because variance ratio can be expressed as a linear combination of the autocorrelation coefficients.

[32] Using previous two or four months gives similar results. Quote-midpoint returns are used to control for bid-ask bounce.

[33] Using q = 2, 3, 4, 6, 8, and 10 gives similar results.

3.1.3 Discussion

Compared to other empirical measures, Hasbrouck's (1993) pricing error has several significant advantages. First, Hasbrouck's measure is free from asset pricing models. As pointed out by Fama (1991), "*…market efficiency per se is not testable. It must be tested jointly with some model of equilibrium, an asset-pricing model*". Proxies that rely on specific asset pricing models thus face potential model-misspecification problem. Such proxies include the Kalman filter in Brennan and Wang (2010) or measures based on momentum profit or post-earnings announcement drift. Hasbrouck's measure, on the other hand, is based only on the random walk hypothesis and does not require a specific model, thus the problem of joint-test is avoided. Second, as suggested by Boehmer and Kelley (2009), Hasbrouck's measure only captures price deviation from random walk caused by uninformed trading. Variance ratio and autocorrelation, however, do not distinguish between information-induced and non-information-induced deviations from random walk. Moreover, Hasbrouck's measure is estimated on a trade-to-trade basis, thus captures all the information involved in trading and puts more weights on periods with active information discovery. Variance ratio and autocorrelation, however, are measured by fixed time interval, leading to information loss and more weights on periods without active price discovery. Therefore, Hasbrouck's measure is expected to give more accurate inferences compared to other empirical proxies.

However, it is important to note that the Hasbrouck's pricing error variance, σ_s^2, is not equivalent to the variance of pricing error, σ_z^2, in the model presented in Section 1. While z is measured based on strong-form efficient price, σ_s^2 only captures mispricing corresponding to weak-form efficient price. In other words, though efficient price should follow a random walk, the implicit random walk process obtained from variance decomposition on intraday data is not necessarily the strong-form efficient price. Meanwhile, as acknowledged by Hasbrouck (1993), low-frequency (daily or longer intervals) temporary components of stock prices, if any, could be mistakenly impounded into the random-walk portion of the VAR decomposition of the intraday price process.

Given the overwhelming evidence of return anomalies in low-frequency data, σ_s^2 is unlikely to be equal to σ_z^2–the latter is probably much greater than former.[34]

The rationale to use σ_s^2/σ_p^2 in the low-frequency setting of this essay comes from the fact that this measure effectively captures non-information-based trading through measuring the deviation of intraday stock prices from random walk (Boehmer and Kelley, 2009). As all kinds of price inefficiency are eventually caused by uninformed trading, σ_s^2/σ_p^2 could be used as a proxy for low-frequency inefficiency, even though it is estimated by intraday prices. To further validate this argument, I test the relation between σ_s^2/σ_p^2 and trading profits based on momentum (Jegadeesh and Titman, 1993) and short-term reversal (Jegadeesh, 1990 and Lehmann, 1990). Greater profits from both trading strategies are generally considered by the literature as indicators of more severe price inefficiency. Specifically, the momentum strategy is implemented as buying last year's winning stocks and selling last year's losing stocks after skipping a month, and the reversal strategy is implemented as buying last week's losing stocks and selling last week's winning stocks after skipping a week to control for microstructure effects. Both strategies are implemented on subsamples of stocks sorted by annual (for momentum strategy) or quarterly (for reversal strategy) average of σ_s^2/σ_p^2. As shown in Table 2.A.1, there are positive and monotonic relationships between σ_s^2/σ_p^2 and trading profits of both strategies. The top 20% stocks sorted by σ_s^2/σ_p^2 generates significant momentum and reversal profits, while the bottom 20% stocks do not exhibit significant profits for either of the strategies. Therefore, though σ_s^2/σ_p^2 is estimated by intraday data, it is positively and strongly associated with low-frequency (weekly, monthly, etc.) price inefficiency and could be used under the monthly setting of this essay.

3.3 Quote-midpoint return

[34] For instance, σ_s^2 on average accounts for 1.29% of return variance in the stock sample of this essay, but French and Roll (1986) find that approximately 4% to 12% of the daily return variance comes from mispricing.

Blume and Stambaugh (1983) argue that, due to Jensen's inequality, bid-ask bounce generates upward biases on stock returns computed with closing prices. This problem could be exacerbated in this essay, as low-price-efficiency stocks tend to have large bid-ask spread. Moreover, Lo and MacKinlay (1990) show that nonsynchronous trading could cause serial correlations, which could distort the variance ratio estimates used in this essay. To control for such microstructure biases, monthly and daily quote-midpoint returns are calculated based on the midpoint of the last quote in regular trading hours of each month or each day. In the case of a stock split or dividend payout in time *t*, the quote midpoint is adjusted accordingly. The monthly quote-midpoint return has an average correlation coefficient of 0.97 with the closing-price return.

3.4 Idiosyncratic volatility

Idiosyncratic volatility ($IVOL$) is estimated by the approach of Ang et al. (2006). For each month, daily stock returns are regressed on the three Fama-French factors,

$$r_{it} = \alpha_{it} + b_{it} r_{mt} + s_{it} SMB_t + h_{it} HML_t + \varepsilon_{it}. \tag{2.11}$$

IVOL is estimated as the product of the residual standard deviation, $\sigma(\varepsilon_i)$, and square root of the number of trading days in that month. Alternatively, following Ferreira and Laux (2007), this essay also uses one minus the *R*-square of the above regression as a relative measure of idiosyncratic volatility (*RIVOL*). In addition, the in-sample estimation of the expected idiosyncratic volatility (*EIV*) of Fu (2009) is also used in cross-sectional regression as a control variable. For each stock-month, the *EIV* is estimated by an exponential-GARCH model based on all available monthly returns up to the end of that month.

3.5 Control variables

Four measures for illiquidity and transaction cost are used in the cross-sectional analysis: 1) trade-weighted relative effective spread (*RES*), estimated as two times the absolute distance between actual transaction price and the prevailing quote midpoint, scaled by the quote midpoint; 2) time-weighted relative quoted spread (*RQS*), estimated as the absolute distance between bid and ask price, scaled by the quote midpoint, 3) Amihud (2002) price impact measure (*Amihud*), estimated by the approach of Acharya and Pedersen (2005), and 4) Liu (2006) measure of non-trading days (*Liu*), estimated as the number of zero-trade days over the previous twelve months adjusted by turnover ratio. *RES* is a preferred measure of transaction cost since it measures the actual costs for traders. However, *RES* may underestimate transaction cost, since trades are relatively infrequent during periods of low liquidity.

Beta, market value, and book-to-market ratio are estimated by the same approach of Fama and French (1992). Market value (*MV*) is the number of shares outstanding multiplied by closing price (CRSP items SHROUT and PRC). *MV* from July to the next June is measured as the market value at the end of June of that year. *BE* at the fiscal year end before last December and *ME* at the end of last December are used to estimate *B/M* from this July to the next June.[35] In the beta estimation, stocks are first assigned to portfolios by their pre-ranking betas and size. Beta estimate of each portfolio is then assigned to stocks over the period when they are in that portfolio.

Analysts following (*ANLY*) measures the total number of analysts that report earnings forecasts for a stock within the past two quarters (IBES item FPI = 6 or FPI = 7). As reported in Table 2.1, stocks have an average of seven analysts following.

Institutional ownership (*IO*) is defined as the total shares held by any institutional investor (who files the 13-f file) at the end of a quarter scaled by the total shares outstanding at the same time. As reported in Table 2.1, nearly half of total shares outstanding of U.S. common stocks are held by institutional investors over 1984-2012.

Price delay (*PD*) is estimated by the same approach of Hou and Moskowitz (2005). At the end of June of each year, two regressions are conducted: 1) weekly stock returns in

[35] Book equity (BE) is estimated by the approach of Gorodnichenko and Weber (2013). It is defined as total shareholders' equity plus deferred taxes and investment tax credit (Compustat item TXDITCQ) minus the book value of preferred stock (Compustat item PSTKQ). Shareholders' equity numbers is taken from Compustat item SEQQ. In case this data is not available, shareholders' equity is computed as sum of common and preferred equity (Compustat items CEQQ and PSTKQ). If neither of the two are available, shareholders' equity is calculated as the differences of total assets and total liabilities (Compustat items ATQ and LTQ).

the prior 12 months are regressed on contemporaneous market return; 2) weekly stock returns in the prior 12 months are regressed on contemporaneous market return and up to four weeks of lagged market returns. *PD* is measured by one minus the ratio of the *R*-square of the first regression to the *R*-square of the second regression. Hou and Moskowitz (2005) consider it to be an inverse measure of the stock's speed of response to market news.

Finally, the probability of informed trading (*PIN*) is estimated on an annual basis by the same approach of Easley, Hvidkjaer, and O'Hara (2002). Based on a structure model of order imbalance, *PIN* measures the probability that a trade is initiated by an informed trader. As reported by Table 2.1, about 17% of all trades in U.S. common stocks over 1984-2012 are initiated by informed traders. This number is very close to the estimates in Easley et al. (2002).

4. Price inefficiency, idiosyncratic volatility, and expected stock return

Price inefficiency could rise through various channels, such as liquidity trading, overreaction or underreaction, price discreteness, temporary price movement by block trading, and index-linked investing. Information asymmetry and illiquidity could exacerbate pricing error and bid-ask spread and transactions costs could limit arbitrages that correct pricing errors. In the empirical analysis, I focus on evaluating and estimating the effect of price inefficiency on stock returns, without directly exploring the reasons for the inefficiency.

This section presents empirical evidence of positive cross-sectional relation between price inefficiency and expected idiosyncratic volatility, and between price inefficiency and expected stock returns. First, Section 3.1 examines empirical determinants of price inefficiency using Fama-MacBeth (1973) cross-sectional regressions. A positive association between inefficiency and future idiosyncratic volatility is then established in Section 3.2 by both time-series and cross-sectional regressions, confirming the model prediction in Section 1. Section 3.3 shows positive cross-section relation between stock returns and lagged inefficiency by Fama-MacBeth regressions. To further quantify the magnitude of this inefficiency premium, Section 3.4 forms quintile portfolios by lagged inefficiency measures and analyzes their risk-adjusted returns. In addition, Section 3.5 investigates the impact of institutional ownership on the inefficiency premium and find supporting evidence to the Merton (1987) predictions.

4.1 Determinants of price inefficiency

Fama-MacBeth (1973) cross-sectional regressions are conducted to examine the empirical determinants of price inefficiency. Specifically, I run the following cross-sectional regression for each of the 347 months in the sample period from February 1984 to December 2012:

$$(\sigma_s^2/\sigma_p^2)_{i,t} = \gamma_{0,t} + \sum_{k=1}^{K} \gamma_{k,t} X_{k,i,t-1} + \varepsilon_{i,t}, \qquad t = 1,2,\dots,T. \tag{2.12}$$

The normalized pricing error variance of Hasbrouck (1993), σ_s^2/σ_p^2, is used as the main proxy for inefficiency. The absolute value of weekly-to-daily variance ratio based on daily quote-midpoint return, $|VR(5)|$, is used as an alternative proxy for robustness test. The control variables, X_k, include market value (*MV*) and book-to-market ratio (*B/M*) estimated by the approach of Fama and French (1992), lagged relative effective spread (*RES*), lagged monthly turnover ratio (*TO*), number of analysts following the stock (*ANLY*), lagged institutional ownership (*IO*), and share price at the beginning of the month. Logarithmic transformation is applied to most variables to control for non-normality. The final coefficient estimates are given by the time-series average of the 347 monthly estimates:

$$\hat{\gamma}_k = \frac{1}{T}\sum_{t=1}^{T} \hat{\gamma}_{k,t}. \tag{2.13}$$

The *t*-statistic is $\hat{\gamma}_k$ scaled by its time-series standard error, which is estimated by the approach of Newey and West (1987) to adjust for potential autocorrelation and heteroskedasticity.

Please insert Table 2.2 here

As presented by Table 2.2, both measures of inefficiency, σ_s^2/σ_p^2 and $|VR(5)|$, increase with *B/M* ratio and effective spread and decrease with stock size, turnover ratio, number of analysts, institutional ownership, and share price. This result is consistent with the literature and intuition that price inefficiency is more likely to occur in small stocks, low-priced stocks, stocks with low liquidity and high transaction cost, and stocks with high information asymmetry.

4.2 Relation between price inefficiency and expected idiosyncratic volatility

I use both time-series regressions and Fama-MacBeth cross-sectional regressions to examine the relation between expected idiosyncratic volatility and inefficiency. I first run the following cross-sectional regression for each of the 347 months in the sample period from February 1984 to December 2012:

$$DV_{i,t} = \gamma_{0,t} + \sum_{k=1}^{K} \gamma_{k,t} X_{k,i,t-1} + \gamma_{K+1,t} (\sigma_s^2/\sigma_p^2)_{i,t-1} + \varepsilon_{i,t}, \qquad t = 1,2,\dots,T. \quad (2.14)$$

The dependent variable, *DV*, is either *IVOL*, the monthly idiosyncratic volatility measured by the approach of Ang et al (2006), or *RIVOL*, the fraction of idiosyncratic volatility in total return volatility. σ_s^2/σ_p^2 is used as the main proxy for inefficiency, while $|VR(5)|$ is used for robustness test. The control variables, X_k, include lagged dependent variable, lagged volatility of log transaction price (σ_P), market value (*MV*) and book-to-market ratio (*B/M*), lagged relative effective spread (*RES*), turnover ratio over the previous 12 months (*TO*), number of analysts following the stock (*ANLY*), and institutional ownership (*IO*). The final coefficient estimates is given by the time-series average of the 347 monthly estimates and the *t*-statistics are based on Newey-West standard errors.

Next, I conduct time-series regressions using the same dependent and independent variables. Specifically, for each stock that has no less than 50 monthly observations over the sample period, I run the following time-series regression:

$$DV_{i,t} = \gamma_{0,i} + \sum_{k=1}^{K} \gamma_{k,i} X_{k,i,t-1} + \gamma_{K+1,i} (\sigma_s^2/\sigma_p^2)_{i,t-1} + \varepsilon_{i,t}, \qquad i = 1,2,\dots,N. \quad (2.15)$$

There are a total of 4,059 stocks included in this regression. The final coefficient estimates is given by the cross-sectional average of the coefficient estimates of all stock:

$$\hat{\gamma}_k = \frac{1}{N}\sum_{i=1}^{N}\hat{\gamma}_{k,i}. \tag{2.16}$$

Please insert Table 2.3 here

Table 2.3 presents the coefficient estimates and *t*-statistics for both cross-sectional and time-series regressions. There are positive cross-sectional and time-series relations between lagged σ_s^2/σ_p^2 and idiosyncratic volatility, either measured by absolute magnitude or by relative fraction in total return volatility. Using variance ratio as proxy of inefficiency gives similar results. Moreover, $IVOL$ is negatively related to size and book-to-market ratio and is positively related to lagged price volatility, illiquidity, and turnover ratio.

4.3 Relation between price inefficiency and expected stock return

The relation between price inefficiency and expected stock return is examined by Fama-MacBeth regressions of monthly stock returns on inefficiency measures and various stock characteristics. Specifically, for each month in the sample period, I run the following regression:

$$R_{i,t} = \gamma_{0,t} + \sum_{k=1}^{K}\gamma_{k,t}X_{k,i,t-1} + \gamma_{K+1,t}(\sigma_s^2/\sigma_p^2)_{i,t-1} + \varepsilon_{i,t}, \qquad t = 1,2,\ldots,T. \tag{2.17}$$

The dependent variable, $R_{i,t}$, is the monthly stock return in excess of risk-free rate. σ_s^2/σ_p^2 is used as the main proxy for inefficiency while $|VR(5)|$ is used for robustness test.

Please insert Table 2.4 here

Table 2.4 reports the time-series averages of coefficient estimates from monthly cross-sectional regressions based on different specifications. In model [1], stock beta, size, B/M, and past return are used as explanatory variables. Stock beta is included to control for cross-sectional difference in exposure to the market risk. Size and B/M are controlled since Fama and French (1992) find that size and B/M are two significant determinants of cross-sectional stock returns. Moreover, compound excess return from month $t-12$ to $t-2$ is included to control for momentum effect found by Jegadeesh and Titman (1993). Consistent with Fama and French (1992), relation between stock return and beta is not statistically significant, while stock return decreases with size and increases with B/M. Consistent with Jegadeesh and Titman (1993), stock return increases with past return. In other words, small stocks, value stocks, and stocks with higher past returns tend to have higher returns.

In model [2] and [3], lagged σ_s^2/σ_p^2 is included as an explanatory variable in addition to stock beta, size, B/M, and past return. Consistent with my hypothesis, there is a positive and significant relation between lagged σ_s^2/σ_p^2 and stock return that is not explained by size, value, or momentum effect.

Model [4] and [5] add lagged *IVOL*, expected idiosyncratic volatility (*EIV*), illiquidity measured by relative effective spread (*RES*), average monthly turnover ratio over the previous 12 months (*TO*), and institutional ownership (*IO*) as additional control variables. Ang et al. (2006) find a negative relation between stock return and lagged *IVOL*. Fu (2009), however, finds a positive relation between stock return and *EIV* estimated by EGARCH approach. It is reasonable to control for both measures as inefficiency could be closely related to idiosyncratic risk. Moreover, the literature generally supports a positive relation between illiquidity and expected return. Controlling for illiquidity, therefore, helps distinguish between the effect of inefficiency and the potential complication from illiquidity premium. Consistent with the literature, stock return decreases with lagged *IVOL* and increases with *EIV* and *RES*. More importantly, the positive relation between lagged σ_s^2/σ_p^2 and stock return is not affected by the inclusion of additional control variables. The robustness of the results is further confirmed by model [6] and [7], where the Amihud (2002) price impact measure (*Amihud*) and Liu (2006) no-trade-day measure (*Liu*) are used as alternative illiquidity proxies. Finally, model [8] and [9] use $|VR(5)|$ as proxy for inefficiency and confirm the robustness of the results.

Results in Table 2.4 provide striking evidence of a positive cross-sectional relation between price inefficiency and average stock return in future. Interestingly, the negative relation between size and stock return almost disappears after the inclusion of σ_s^2/σ_p^2 as an explanatory variable, indicating the possibility that the size effect may be a manifestation of the inefficiency premium.

4.4 Returns of portfolios formed by price inefficiency

After confirming the positive relation between price inefficiency and average stock returns by cross-section regressions, I form portfolios by inefficiency to conduct a more intuitive and easy-to-interpret examination about the magnitude of the inefficiency premium. At the beginning of each month, all stocks in the sample are sorted into five value-weighted portfolios with equal number of stocks by their lagged price inefficiency measures. The High (Low) quintile contains the 20% of stocks with the greatest (lowest) inefficiency in the previous month. Portfolios are weighted by stock size at the beginning of the month and are rebalanced monthly. Time series of portfolio returns are obtained by concatenating the monthly value-weighted average returns of each portfolio and are then evaluated by a time-series regression on empirical risk factors. The time-series averages of the excess returns (r), *FF3*-adjusted returns (α_{FF3}), and *FF3*-and-momentum-adjusted returns (α_{FF4}) are estimated for each quintile portfolio as well as a hypothetical zero-investment portfolio that is long in the High quintile and short in the Low quintile. Equally-weighted four-factor adjusted return ($\alpha_{FF4,EW}$) is also reported. To reduce potential microstructure influences (bid-ask bounce and nonsynchronous trading, etc.), quote-midpoint returns (r^{MP}) are used, in addition to closing-price returns, to estimate monthly portfolio returns.

Please insert Table 2.5 here

Panel A of Table 2.5 reports returns of portfolios formed by lagged σ_s^2/σ_p^2. Consistent with the hypothesis and the results from cross-sectional regressions, stocks in the High quintile significantly outperform stocks in the Low quintiles. The zero-investment portfolio generates monthly raw returns of 0.40% (t = 2.12) and four-factor-adjusted returns of 0.50 (t = 3.95). When quote-midpoint returns are used to replace closing-price returns, the zero-investment portfolio generates even greater raw returns and four-factor-adjusted returns, 0.43% (t = 2.27) and 0.52% (t = 4.17), respectively. Whether this outperformance can be realized by investors, however, depends on transaction costs and is analyzed in Sections 5.

Panel B of Table 2.5 presents time-series averages of value-weight stock characteristics of each of the quintile portfolios in each month, as well as the difference between the characteristics of the High and Low quintiles. Consistent with intuition and the results of the cross-sectional regressions in Table 2.2, stocks with greater price inefficiency tend to have smaller size, higher B/M ratio, higher IVOL, larger effective spread, lower turnover ratio, fewer analyst following, and lower institutional ownership. It is worth noting that the difference between the IVOL of the High quintile and the Low quintile at the present month, 0.037, is noticeably larger than the difference in the previous month, 0.014. In other words, while the average IVOL of stocks in the High quintile is 15% higher than that in the Low quintile in the previous month, it becomes 43% higher in the present month. This indicates that the larger IVOL of stocks in the High quintile is not merely a result of time-persistence in IVOL, but probably comes more from price inefficiency in the previous month. It is intuitive evidence to the hypothesis of this essay that price inefficiency increases expected IVOL.

There could be a potential concern about the magnitude of the increase in IVOL. Though an increase of 28% (= 43% - 15%) in IVOL is not tiny, the cross-sectional dispersion in IVOL is much large than this magnitude. For example, as shown in Table 2.1, the 25^{th} percentile of IVOL is 0.055, while the 75^{th} percentile increases noticeably to 0.134, or an increase of 142%. Why could the relatively limited increase in IVOL induced by price inefficiency (28%) make a noticeable difference in stock returns? One reasonable explanation is that such an increase represents a change in the *expected* IVOL, which is much more closely related to investors' investment decision making compared with the *realized* IVOL in the previous month.

Panel C of Table 2.5 exhibits returns of portfolios formed by absolute value of lagged variance ratio. Again, the High quintile significantly outperforms the Low quintile by 0.36% (t = 2.16) in raw return and 0.37% (t = 2.09) in four-factor-adjusted return. Inferences based on quote-midpoint returns are quite similar.

4.5 Role of shareholders' portfolio diversification

This essay relies on a framework where under-diversification prompts investors to demand return premium for bearing price inefficiency. This implies that stocks with higher fractions of shares held by well-diversified investors, typically the institutional investors, should exhibit less pronounced inefficiency premium. To test this hypothesis, at the beginning of each month, I sort all stocks into 20 portfolios by their lagged σ_s^2/σ_p^2 and lagged *IO* independently and examine the return dispersion between the high and the low price inefficiency quintile within each *IO* quartile. Consistent with the hypothesis, the High-Low dispersion becomes much weaker (0.21%) and is not statistically significant (t = 1.22) in stocks with the highest IO. Results based on only NYSE stocks are qualitatively the same. For a sub-sample of Russell 2000 stocks which probably have much greater price inefficiency, the High-Low dispersion is still statistically significant in stocks with the highest *IO*. However, there is a large and monotonic decline in the magnitude of this dispersion from low *IO* stocks (1.65%, t = 4.97) to high *IO* stocks (0.60%, t = 2.43).

Please insert Table 2.6 here

4.6 Discussion

The literature of idiosyncratic risk premium has been troubled in recent years by the puzzling finding of a negative relation between stock returns and lagged IVOL. Why does portfolios based on price inefficiency in this essay give different inference compared to portfolios based on lagged IVOL that are commonly used in the literature? Studies in recently years have pointed out several possible complications associated with lagged IVOL that could lead to a false negative relation between stock returns and idiosyncratic risk. First, unexpected idiosyncratic volatility is followed by return reversals. Huang et al. (2010) find that the negative relation between lagged idiosyncratic volatility and future returns disappears after controlling for return reversals and that there is a positive relation between condition IVOL and future returns. Similarly, Chua, Goh, and Zhang (2010) find that high IVOL is associated with high unexpected IVOL (UIV), which leads to high contemporaneous stock returns thus low future returns. After controlling for UIV, expected idiosyncratic risk is positively associated with expected returns. Second, lagged IVOL could be inversely related to certain risk factors that are missed in traditional asset pricing models. Chen and Petkova (2012) propose that portfolios with high IVOL have positive exposures to innovations in average stock variance and thus lower expected returns. Moreover, lagged IVOL is probably an inappropriate idiosyncratic risk proxy due to autocorrelation in IVOL. Using an EGARCH approach to estimate the expected idiosyncratic volatility (EIV), Fu (2009) finds a positive relation between future returns and EIV. In sum, sorting stocks into portfolio by lagged IVOL could introduce complications that conceal the true relation between idiosyncratic risk and expected returns. To obtain accurate inferences, we need to use proxies of *expected* IVOL rather than *realized* IVOL, as well as to properly control for potential complications associated with lagged IVOL.

An appealing feature of this essay is that the long position (relatively inefficient stocks) and the short position (relatively efficient stocks) do not differ significantly in their lagged IVOL, thus will largely alleviate the complications mentioned above. More importantly, as current inefficiency produces future IVOL, stocks with greater inefficiency have higher expected IVOL in future, even though they do not have significantly greater present IVOL. Therefore, under the setting of this essay, I am able to compare the performance of two groups of stocks with similar lagged IVOL but different expected IVOL in future. This probably delivers more reliable inference about the true relation between idiosyncratic risk and expected returns.

Arguments of this essay rely on an assumption that common investors are able to detect price inefficiency and make investment decision accordingly. Theoretically, uninformed investors should not be able to precisely detect the presence, magnitude and direction of pricing error; otherwise, price inefficiency will be corrected immediately. Practically, however, it is possible for investors to have a rough perception about the existence and severity of price inefficiency, though they are unlikely to know the directions of the pricing errors. For example, investors could infer the presence and severity of inefficiency from the magnitude of the deviation of stock price from random walk or from the severity of market anomalies such as post earnings announcement drift. Inefficiency could be greater when there is large dispersion in analyst forecasts or when there is noticeable disagreement in investors' opinions about the fundamentals. Stocks with greater information asymmetry, such as fewer analysts or more insider holdings, probably have more severe inefficiency. Stocks with larger proportion of retail investors and smaller proportion of institutional investors are also more likely to be mispriced. Therefore, it is reasonable to believe the ability of common investors to generate a rough perception of the existence and relative severity of price inefficiency and to make investment decisions accordingly.

5. Robustness tests

Price inefficiency is difficult to measure and is usually entangled with small size, illiquidity, high transaction cost, low investor recognition, and high information risk. Thus this section provides extensive tests about other possible explanations for the positive relation between price inefficiency and expected stock returns. Specifically, I confirm the robustness of the main finding of this essay to potential size effect, illiquidity, transaction cost, lagged idiosyncratic volatility, January effect, and information risk. Moreover, the result is also robust to alternative specifications such as using NYSE breakpoints or market share breakpoints to form portfolios, skipping one month between the ranking period and the holding period, different sub-periods, etc. In addition, the inefficiency premium found by this essay is clearly different from the "mispricing return premium" proposed by Brennan and Wang (2010).

Please insert Table 2.7 here

5.1 Control for stock characteristics

5.1.1 Stock size

As the earlier sections of this essay have shown a negative relation between inefficiency and stock size, it is possible that the inefficiency premium is merely a result of the size effect. I adopt two approaches to disentangle the size effect from the inefficiency premium. First, within each size quartile based on NYSE size breakpoints, stocks are sorted into value-weighted inefficiency quintiles and their four-factor-adjusted returns are examined. If inefficiency premium comes from the size effect, outperformance of high σ_s^2/σ_p^2 stocks should be largely reduced within each size quartile. As shown in Panel A of Table 2.7, however, return spreads between high- and low-inefficiency stocks are positive and significant in each of the size quartiles.

Sorting stocks based on size, however, may cause systematic bias, as underpriced stocks are more likely to appear in small-size quartiles while overpriced stocks are more likely to enter large-size quartiles. To avoid this problem, I repeat the tests based on constituents of certain stock indexes. As stocks are not frequently added into or removed from an index, temporary price movements are not likely to cause systematic bias in the classification of stock size. Specifically, I use S&P 500 constituents to represent large-cap stocks, use Russell 1000 constituents as a sample of large- and mid-cap stocks, and use Russell 2000 constituents to represent small-cap stocks. I also repeat the test on NYSE-listed stocks, and on all stocks in NYSE/AMEX/NASDAQ, including the ones with the smallest market capitalizations. High-inefficiency stocks significantly outperform low-inefficiency stocks in all of these settings, as reported in Panel A of Table 2.7. Therefore, the inefficiency premium is robust to the size effect.

5.1.2 Illiquidity/transaction costs

There is a positive cross-sectional correlation between price inefficiency and illiquidity measured by effective spread (not reported), due to several potential reasons. First, noise traders who typically provide liquidity to the market may be reluctant to trade when facing large information asymmetry. Second, as market makers are facing higher probability to trade with informed traders when stocks have greater price inefficiency, they are likely to widen bid-ask spreads to compensate for their expected loss (Copeland and Gailai, 1983). Third, inefficiency could be a result of illiquidity, as illiquidity increases the cost of arbitrage. As the literature generally supports a positive relation between illiquidity and expected returns[36], controlling for illiquidity helps distinguishing between the inefficiency premium and the illiquidity premium.

I adopt several approaches to control for illiquidity. First, I sort stocks into subsamples based on their lagged *RES* or *Amihud*, and then sort stocks within each subsample into quintiles by σ_s^2/σ_p^2. Second, I generate four subsamples: *i*) excluding stocks with share prices lower than $5 at the beginning of the month, *ii*) large/liquid stocks that are among the top half by size and bottom half by *RES*, *iii*) large/high-priced stocks that are among the top half by size and top half by share price, and *iv*) liquid/high-priced stocks that are among the bottom half by *RES* and the top half by share price. Third, I add either the Pastor and Stambaugh (2003) or the Liu (2006) traded liquidity factor into the traditional four-factor model to obtain the five-factor-adjusted returns of portfolios classified by σ_s^2/σ_p^2. If the abnormal returns associated with price inefficiency come from the liquidity premium, they should disappear or largely decline: 1) within each of the subsamples by illiquidity, and/or 2) after controlling for traded liquidity factors.

Panel B of Table 2.7 presents the results of the tests above. Contrary to the illiquidity explanation, stocks with greater price inefficiency continue to significantly outperform stocks with low price inefficiency in all of the subsamples. Neither does the inclusion of traded liquidity factors reduce the magnitudes and significances of the outperformance. Furthermore, as shown in Table 2.3, inclusion of illiquidity measures of *RES*, *Amihud*, or *Liu* does not erode the significance of inefficiency in explaining the cross-sectional variation of stock returns. Therefore, the inefficiency premium is not likely to be a manifestation of liquidity premium.

5.1.3 Lagged idiosyncratic volatility

Some recent studies have documented a negative relation between lagged IVOL and future stock return. As the main proxy for price inefficiency in this essay, σ_s^2/σ_p^2, is

[36] Amihud and Mendelson (1986), Brennan and Subrahmanyam (1996), Chordia, Subrahmanyam, and Anshuman (2001), Amihud (2002), and Pastor and Stambaugh (2003), among many others.

the pricing error variance normalized by total return volatility, it is natural to concern about the impact of an IVOL shock on σ_s^2/σ_p^2. If σ_s^2/σ_p^2 is lowered due to such a shock, there could be a positive relation between σ_s^2/σ_p^2 and future stock returns. However, this is unlikely to be the case, as the average cross-sectional correlation between σ_s^2/σ_p^2 and IVOL is positive (not reported). To further confirm the robustness of the inefficiency premium to lagged IVOL, I generate four subsamples based on lagged IVOL and examine the performance of portfolios formed by σ_s^2/σ_p^2 within each subsample. As reported in Panel C of Table 2.7, contrary to the IVOL explanation, the return premium for price inefficiency continues to exist in three of the four subsamples. Therefore, it is unlikely for lagged IVOL to be an explanation.

It is, however, interesting to discuss the reason why the return spread is insignificant in the high IVOL subsample. There could be two possible explanations. First, the return reversals following IVOL shock may be large enough to override the inefficiency premium, leaving no significant relation between price inefficiency and average stock returns in the subsequent period. Second, high IVOL may obstruct investors' ability of detecting price inefficiency. I leave this topic for future studies.

5.1.4 Price delay

Hou and Moskowitz (2005) propose a measure of the delay of stock price (*PD*) in response to market movements. They find that the most delayed firms earn a large return premium unexplained by size, liquidity, or microstructure effects. They attribute this finding to frictions associated with investor recognition in the Merton (1987) framework. As delayed firms are small firms with large frictions, there is probably an overlap between the most delayed stocks and the most mispriced stocks, resulting in a positive relation between price inefficiency and future stock returns. To preclude this possibility, I sort stocks into quartiles by *PD* and examine the performance of portfolios formed by σ_s^2/σ_p^2 within each *PD* quartiles. In contrast to the price delay explanation, inefficiency still generates return premiums in all *PD* quartiles (Panel D in Table 2.7).

5.1.5 Information risk

Easley, Hvidkjaer, and O'Hara (2002) find that stocks with high probability of informed trading (*PIN*) are associated with high expected return. As informed investors always overweight on underpriced stocks and underweight on overpriced stocks, uninformed investors are always on the wrong side, even with considerable level of diversification. Therefore, common investors demand compensation for holding stocks with high probability of informed trading. As price inefficiency should be strongly associated with informed trading, it is naturally a concern about whether the inefficiency premium comes from the *PIN* premium.

This argument, however, may not be a real challenge to the findings of this essay. The model proposed by Hughes, Liu, and Liu (2007) show that effect of asymmetric information on expected returns is diversifiable in a large economy. By decomposing the asymmetric information portion and liquidity portion of *PIN*, Duarte and Young (2010) find that only the latter is priced. In other words, both theoretical and empirical evidence in recent studies suggest that information risk is not associated with return premium. To further confirm the robustness of my findings, I sort stocks into quartiles by *PIN* and examine the performance of portfolios formed by σ_s^2/σ_p^2 within each *PIN* quartiles. As expected, inefficiency premium exists in all *PIN* quartiles (Panel E in Table 2.7).

5.1.6 Alternative specifications

Alternative specifications are also used to test the robustness of the results. First, the original Hasbrouck's pricing error variance (σ_s^2) is used to replace the scaled variant σ_s^2/σ_p^2 as an alternative measure of inefficiency. To make the cross-sectional comparison of σ_s^2 meaningful, I first sort stocks into deciles by their intraday (log) price variance (σ_p^2) to control for the cross-sectional difference in total return volatility[37]. Within each decile by σ_p^2, I sort stocks into quintiles by their σ_s^2. The VW returns of each σ_s^2 quintile are estimated based on all stocks in the same σ_s^2 quintile across different σ_p^2. Second, I apply

[37] Without this conditional double sort, forming portfolios solely by σ_s will be largely the same as forming portfolios by the (log) price volatility, as the estimation of σ_s is significantly affected by σ_p.

the following settings to the tests: *i*) use NYSE breakpoints of σ_s^2/σ_p^2, *ii*) use market share breakpoints, *iii*) skip one month between the ranking period and the holding period to control for potential microstructure effects, *iv*) drop the top 10% or 20% stocks by σ_s^2/σ_p^2 to examine whether the inefficiency premium is a pervasive phenomenon or only concentrates in a small number of stocks with the greatest inefficiency, *v*) subperiods of pre-decimalization (1984-2000) and post-decimalization (2001-2012), as well as exclude the global financial crisis (2008-09), *vi*) drop January observations to isolate potential January effect. The positive relation between stock returns and lagged inefficiency is robust to all these different specifications. Finally, all tests in Table 2.7 are repeated using quote-midpoint returns instead of closing-price returns and the results are as pronounced as before.

5.2 Mispricing return premium of Brennan and Wang (2010)

Blume and Stambaugh (1983) and Brennan and Wang (2010) propose a mispricing return premium generated from Jensen's inequality. In short, as price is a non-linear function of expected return, unconditional rational price with random error will generate upward bias to the expected return. This can be illustrated by a simple numeric example. Suppose the effective price for a stock is always \$10 but the market price is subject to either \$1 or -\$1 pricing error with equal probability at time *t*. Further suppose that this pricing error will be corrected at time *t*+1. The unconditional expected price of the stock at time t is \$10 as the pricing error has expected value of zero. The expected rate of return from time *t* to *t*+1, however, is 1.01% rather than zero. Formally, define $\tilde{Z}$ is a random pricing error coefficient satisfying $P \equiv P^*\tilde{Z}$, where P is the market price and $P*$ is the efficient price. The expected return of mispriced stock could be expressed by:

$$E[1+\tilde{R}] = (1+r^*)E\left[\frac{1}{\tilde{Z}}\right], \qquad (2.18)$$

where r^* is the expected return of the efficient price. Unconditional rational price implies $E[\tilde{Z}] = 1$, so that $E[1/\tilde{Z}] > 1$ when $Var(\tilde{Z}) > 0$. Therefore,

$$E[1+\tilde{R}] > 1+r^*. \qquad (2.19)$$

Jensen's inequality, however, is not able to explain the main findings of this essay, because it has little impact on value-weighted portfolios used in this essay. Expected return of an EW portfolio is the arithmetic average of the expected returns of its component stocks. When the expected returns of individual stocks are upward biased, so is the expected portfolio return. Expected return of a VW portfolio, however, is only subject to upward biased produced by pricing errors at the *portfolio* level. As random pricing errors of individual stocks are mostly diversified away in a large portfolio, errors in the market value of that portfolio will have fairly small magnitude, leaving only a tiny upward bias on expected portfolio returns[38]. In an ideal setting of perfect diversification, market value of a VW portfolio will always be its intrinsic value, so that the portfolio return will always be unconditionally rational. In sum, while Jensen's inequality mechanically produces upward biases on expected *rates of return* of mispriced stocks, findings of this essay represents greater expected *values* of mispriced stocks induced by rational investors' risk aversion.

Please insert Table 2.8 here

I use simulation to further confirm this argument. Based on the framework presented in Section 1, I simulate 500 series of stock prices with random pricing errors across 348 periods for 100 iterations. Each stock is assumed to have only one share so its market value is equal to its share price. Market values of all S&P 500 stocks at the end of 2011 are assigned to the simulated stocks to mimic the realistic cross-sectional variation in stock size. To simplify the simulation process and to eliminate complications from factors other than mispricing, I assume that all stocks always earning zero return by their efficient prices. Pricing errors have zero mean, zero autocorrelation, and homogeneous standard deviations across stocks and over time. Specifically, various pricing error volatility are tested: σ_z = 0.5%, 1%, 2%, 3%, 5%, 10%, 15%, and 20%. I then compare returns of EW and VW portfolios based on the simulated prices. Panel A of Table 2.8 confirms Brennan and Wang's (2010) argument: mispricing generates positive return premium and its magnitude increases with the magnitude of pricing error measured by σ_z. More

[38] Blume and Stambaugh (1983) have a similar argument that the upward biases from bid-ask bounce can be largely eliminated by the use of a "buy-and-hold" strategy, which shares a similar concept as a VW portfolio. Specifically, buy-and-hold strategy does not use regular rebalancing. As a result, returns to a buy-and-hold strategy will be largely free from the impact of Jensen's inequality.

importantly, the mispricing return premium induced by Jensen's inequality has little effect on the returns of VW portfolios. Returns of the VW portfolio is always a tiny portion of those of the EW portfolios and is far from statistically significant. For instance, even when an extremely large mispricing ($\sigma_z = 20\%$) is assumed in all the 500 stocks, VW portfolio only has a return of 0.03% (t = 0.57), far lower than the 4.07% (t = 112.95) return of the EW portfolio. The results are similar even when only 100 stocks are included into the portfolio[39] (Panel B of Table 2.8). Therefore, the upward return bias proposed by Blume and Stambaugh (1983) and Brennan and Wang (2010) is not able to explain the main findings of this essay.

6. Extensions

6.1 Net return of trading strategies based on price inefficiency

Robustness tests in the last section show that the inefficiency premium is not a result of cross-section variations in transaction costs. This section further explores whether trading strategies based on inefficiency could realize profits after deducting transaction cost. Specifically, I construct zero-investment portfolios that are long in the top and short in the bottom stocks sorted by lagged price inefficiency, and hold the portfolios for either one month, one quarter, half year, or one year. The portfolio net returns are estimated as the value-weighted quote-midpoint returns of all its component stocks minus the percentage transaction costs incurred by each rebalancing event[40].

---

[39] Fewer stocks in the portfolio means less diversification, so that pricing error at the portfolio level will have higher magnitude thus generates greater impact on the portfolio return. The number of 100 is chosen because, in my empirical analysis, the lowest number of stocks in the portfolios is 100 (i.e. quintile portfolios based on S&P 500 stocks).

[40] The transaction cost is estimated as follows. First, I calculate the absolute value of holding change of each component stock. For example, if a stock accounts for 5% of the total asset of the portfolio before the rebalancing and 3% after that, I record a 2% change in the holding of this stock. Second, the incurred transaction cost is estimated as the holding change multiplied by half of the average *RES* (or *RQS*) of this stock during the rebalancing month. For example, given the *RES* to be 1%, transaction cost incurred by this holding change equals to 2%×1% = 0.02% of the total asset value of the portfolio at the end of that month. Finally, the transaction cost of this rebalancing event is calculated as the sum of the transaction costs of all its component stocks. For example, if all stocks have an average *RES* of 1%, the total transaction costs of this rebalancing event will be 1% of its total asset value immediately before this rebalancing.

Either relative effective spread (*RES*) or relative quoted spread (*RQS*) during the contemporaneous month of the rebalancing is used to measure round-trip transaction costs. Both measures are widely used in the literature but they differ significantly in magnitudes. For example, Petersen and Fialkowski (1994) find that effective spreads are only half of the quoted spreads and that increase in the quoted spread results in only 10%-22% increase in the effective spreads. The quoted spread significantly overestimates the true transaction costs because many trades are executed inside the quoted spreads. Therefore, *RES* is used as the main proxy for transaction costs while *RQS* is only for robustness.

Please insert Table 2.9 here

Table 2.9 presents the net returns of zero-investment portfolios based on σ_s^2/σ_p^2. When *RES* is used to measure transaction cost, the zero-investment portfolio that is long in the top 20% and short in the bottom 20% stocks sorted by lagged σ_s^2/σ_p^2 generates monthly four-factor-adjusted return of 0.41% (t = 2.20) when being rebalanced quarterly. The strategy based only on S&P 500 stocks produces monthly abnormal return of 0.28% (t = 1.71) with semiannual rebalancing. Strategies based on Russell 2000 stocks provide monthly abnormal return as high as 0.77% (t = 3.93) with annual rebalancing. When *RQS* is used to measure transaction cost, the abnormal returns are noticeably lowered but could still be positive for certain strategies. For example, the monthly abnormal return is 0.21% (t = 1.71) with annual rebalancing when buying the top 20% and short-selling the bottom 20% stocks sorted by lagged σ_s^2/σ_p^2, or it could be 0.64% (t = 3.18) if implemented only on Russell 2000 stocks.

Of course, effective spread of quoted spread may not fully capture all possible transaction costs. For example, price movements induced by block trades could be partially missed by either *RES* or *RQS*. Other complications such as short-sell constraints are also ignored in analysis above. Nevertheless, result in Table 2.9 still provides further support to the robustness of the inefficiency premium. It also sheds light on trade strategies that could potentially benefit large institutional investors who are more tolerant to random pricing errors in individual stocks.

6.2 Two-factor-adjusted returns of portfolios classified by size, B/M, and momentum

It will be interesting to examine whether the inefficiency premium contribute to traditional risk factors such as the size, value, and momentum effects. In the formal tests in Table 2.10, I construct the inefficiency premium factor (*IP*) as the dispersion between the value-weighed returns of the top 20% and the bottom 20% stocks in my sample sorted by σ_s^2/σ_p^2. A two-factor model based on the market risk factor (*MKTRF*) and the *IP* is then used to test the robustness of the size, value, and momentum effects.

Please insert Table 2.10 here

In Panel A of Table 2.10, all stocks are sorted into size quintiles by a similar approach of Fama and French (1993). Specifically, at the beginning of each month, stocks are sorted into 15 (5×3) value-weighted portfolios by their *MV* and *B/M* ratios independently. Returns of each size quintile are then computed as the arithmetic average of the returns of the corresponding size quintiles in value, neutral, and growth stocks. As reported in Table 2.10, small stocks outperform big ones by 0.76% (t = 2.23) in raw returns and 0.61% (t = 1.88) in market-adjusted returns. The two-factor-adjusted return (α_2), however, becomes negative (-0.07%) and statistically insignificant (t = -0.25). Regression coefficient of the *MRP* is 1.12 and highly significant (t = 12.66). Combined with the evidence that small stocks exhibit greater price inefficiency, the disappearance of the size effect after controlling for the inefficiency premium probably indicates that the size effect is a manifestation of the inefficiency premium.

Panel B of Table 2.10 conducts a similar test on the value effect. At the beginning of each month, stocks are sorted into 10 (5×2) value-weighted portfolios by their *B/M* ratios and *MV* independently. Returns of each *B/M* quintile is then computed as the arithmetic average of the returns of the corresponding *B/M* quintiles in small stocks and in large stocks. As reported in Table 2.10, value stocks outperform growth stocks by 0.79% (t = 3.12) in raw returns and 0.90% (t = 3.32) in market-adjusted returns. Value stocks also exhibit much greater pricing errors, as average σ_s^2/σ_p^2 in the top *B/M* quintile is 0.88% while that in the bottom *B/M* quintile is only 0.29%. However, after the inefficiency premium being adjusted, the value stocks still significantly outperform the growth stocks

(0.80%; t = 3.99). Coefficient of the *MRP* is positive but has small magnitude (0.15; t = 2.31). Therefore, inefficiency premium only has very limited explanatory power on the value effect.

Finally, Panel C of Table 2.10 conducts a similar test on the momentum effect. At the beginning of each month, stocks are sorted into 5 value-weighted portfolios by their compounding returns from month t-12 to t-2. As reported in Table 2.10, past winner stocks outperform past loser stocks by 0.71% (t = 1.69) in raw returns and 0.90% (t = 2.52) in market-adjusted returns. Consistent with the hypothesis of this essay that price inefficiency produces contemporaneous price discount (thus higher expected return), past winner stocks exhibit much less pricing errors than past loser stocks. As a result, the outperformance of past winner stocks becomes even stronger (1.18%; t = 2.93) after the inefficiency premium being controlled, and the *MRP* has a negative coefficient of -0.35 (t = -2.58).

In sum, the return premium for price inefficiency entirely explains the size effect but has very limited contribution to the value effect. The momentum effect, on the other hand, is inversely related to the inefficiency premium.

7. Conclusion

Based on a sample of U.S. common stocks listed in NYSE, AMEX, and NASDAQ from 1984 to 2012, this essay illustrates a positive cross-sectional relation between expected stock returns and price inefficiency, measured by deviations of stock prices from the random walk. Specifically, the dispersion of annualized four-factor-adjusted returns between the top 20% stocks and the bottom 20% sorted by lagged price inefficiency is 6.15%. This outperformance entirely explains the size effect, but is not explained by size, momentum, illiquidity, transaction cost, lagged idiosyncratic volatility, information risk, or the return premium produced by Jensen's inequality. It is more pronounced in stocks owned more by retail investors, but declines remarkably in stocks owned more by institutional investors. It is consistent with the hypothesis that underdiversified investors demand return premiums for bearing extra idiosyncratic risk induced by price inefficiency.

This essay is related to the existing literature, including Blume and Stambaugh (1983) and Brennan and Wang (2010), about positive cross-sectional relations between price inefficiency and expected returns, but has a completely different concept. It illustrates that price inefficiency affects expected returns not only via an upward bias mechanically produced by Jensen's inequality, but also via the investment decisions of rational, underdiversified investors. Moreover, this essay provides new supporting evidence to the notion that idiosyncratic risk is priced. Many recent studies about idiosyncratic risk suggest that lagged idiosyncratic volatility is not a proper measure of expected idiosyncratic volatility. It could also bring complications such as return reversals or missing-factor problem, distorting the observed relation between average stock return and idiosyncratic risk. From a unique prospective that price inefficiency could induce extra idiosyncratic volatility in future, this essay is able to largely avoid such complications and focus on the relation between expected idiosyncratic volatility and stock return. The return premium associated with low-price-efficiency stock is clearly evidence that idiosyncratic risk is priced.

References 2

Acharya, Viral V., and Lasse Heje Pedersen, 2005. Asset pricing with liquidity risk. *Journal of Financial Economics* 77, 375-410.

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

Amihud, Yahov, and Haim Mendelson, 1986. Asset pricing and the bid-ask spread. *Journal of Financial Economics* 17, 223-249.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006. The cross-section of volatility and expected returns. *Journal of Finance* Vol. 61, Issue 1, 259-299.

Arnott, Robert, Jason Hsu, Jun Liu, and Harry Markowitz, 2006. Does noise create the size and value effects? Unpublished working paper.

Barber, Brad M. and Terrance Odean, 2000. Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance* Vol. LV, Issue 2, 773-806.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998. A model of investor sentiment. *Journal of Financial Economics* Vol 49, No. 3, 307-343.

Black, Fischer, 1986. Noise. *Journal of Finance* Vol. XLI, Issue 3, 529-543.

Blume, Marshall E. and Robert F. Stambaugh, 1983. Biases in computed returns. *Journal of Financial Economics* 12, 387-404.

Boehmer, Ekkehart, and Eric K. Kelley, 2009. Institutional investors and the informational efficiency of prices. *Review of Financial Studies* 22 n 9, 3563-3594.

Boehmer, Ekkehart and Juan Wu, 2012. Short selling and the price discovery process. *Review of Financial Studies* Vol. 26, No. 2, 287-322.

Boehmer, Ekkehart, Gideon Saar, and Lei Yu, 2005. Lifting the Veil: An analysis of pre-trade transparency at the NYSE. *Journal of Finance* Vol. LX, No. 2, 783-815.

Brennan, Michael J. and Ashley W. Wang, 2010. The mispricing return premium. *Review of Financial Studies* Vol. 23, No. 9, 3437-3468.

Carhart, Mark M., 1997. On persistence in mutual fund performance. *Journal of Finance* Vol. LII, No. 1, 57-82.

Chen, Zhanhui and Ralitsa Petkova, 2012. Does idiosyncratic volatility proxy for risk exposure? *Review of Financial Studies* Vol. 25, No. 9, 2745-2787.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2008. Liquidity and market efficiency. *Journal of Financial Economics* 87, 249-268.

Chua, Choong Tze, Jeremy Goh, and Zhe Zhang, 2010. Expected volatility, unexpected volatility, and the cross-section of stock returns. *Journal of Financial Research* Vol. XXXIII, No. 2, 103-123.

De Bondt, Werner F. and Richard Thaler, 1985. Does the stock market overreact? *Journal of Finance* Vol. 40, No. 3 793-805.

Duarte, Jefferson and Lance Young, 2010. Why is *PIN* priced? *Journal of Financial Economics* 91, 119-138.

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara, 2002. Is information risk a determinant of asset returns? *Journal of Finance* Vol. LVII, No. 5, 2185-2222.

Fama, Eugene F., 1991. Efficient Capital Markets: II. *Journal of Finance* Vol. XLVI, No. 5, 1575-1617.

Fama, Eugene F. and Kenneth R. French, 1992. The cross-section of expected stock returns. *Journal of Finance* Vol. XLVII, No. 2, 427-465.

Fama, Eugene F. and Kenneth R. French, 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* Vol. 33, Issue 1, 3-56.

Fama, Eugene F. and James D. MacBeth, 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* Vol. 81, No. 3, 607-636.

Ferreira, Miguel A. and Paul A. Laux, 2007. Corporate governance, idiosyncratic risk, and information flow. *Journal of Finance* Vol. LXII, No. 2, 951-989.

French, Kenneth R. and Richard Roll, 1986. Stock return variances: The arrival of information and the reaction of traders. *Journal of Financial Economics* 17, 5-26.

Fu, Fangjian, 2009. Idiosyncratic risk and the cross-section of expected stock returns. *Journal of Financial Economics* 91, 24-37.

Glosten, Lawrence R., and Paul R. Milgrom, 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* Vol. 14, Issue 1, 71-100.

Goetzmann William N. and Alok Kumar, 2008. Equity portfolio diversification. *Review of Finance* 12, 433-463.

Gorodnichenko, Yuriy and Michael Weber, 2013. Are sticky prices costly? Evidence from the stock market. Unpublished working paper.

Griffin, John M., Patrick J. Kelley, and Federico Nardari, 2010. Do market efficiency measures yield correct inferences? A comparison of developed and emerging markets. *Review of Financial Studies* Vol. 23, No. 8, 3225-3277.

Grossman, Sanford J., and Joseph E. Stiglitz, 1980. On the impossibility of informationally efficient markets. *American Economic Review*, Vol. 70, No. 3, 393-408.

Hasbrouck, Joel, 1993. Assessing the quality of a security market: A new approach to transaction-cost measurement. *Review of Financial Studies* Vol. 6, No. 1, 191-212.

Hotchkiss, Edith S. and Tavy Ronen, 2002. The informational efficiency of the corporate bond market: An intraday analysis. *Review of Financial Studies* Vol. 15, No. 5, 1325-1354.

Hou, Kewei and Tobias J. Moskowitz, 2005. Market frictions, price delay, and the cross-section of expected return. *Review of Financial Studies* Vol. 18, No. 3, 981-1020.

Huang, Wei, Qianqiu Liu, S. Ghoo Rhee, and Liang Zhang, 2010. Return reversals, idiosyncratic risk, and expected returns. *Review of Financial Studies* Vol. 23, No. 1, 147-168.

Hughes, John S., Jing Liu, and Jun Liu, 2007. Information asymmetry, diversification, and cost of capital. *Account Review* Vol. 82, No. 3, 705-729.

Jegadeesh, Narasimhan and Sheridan Titman, 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* Vol. 48, No. 1, 65-91.

Kumar, Alok, 2009. Who gambles in the stock market? *Journal of Finance* Vol. 64 No.4, 1889-1933.

Kyle, Albert S., 1985. Continuous auctions and insider trading. *Econometrica* Vol. 53, No. 6, 1315-1335.

Lee, Charles M. C., and Mark J. Ready, 1991. Inferring trade direction from intraday data. *Journal of Finance* Vol. 46, Issue 2, 733-746.

Levy, Haim, 1978. Equilibrium in an imperfect market: A constraint on the number of securities in the portfolio. *American Economic Review* Vol. 68, No. 4, 643-658.

Liu, Weimin, 2006. A liquidity-augmented capital asset pricing model. *Journal of Financial Economics* 82, 631-671.

Lo, Andrew W., and A. Craig MacKinlay, 1988. Stock market prices do not follow random walks: evidence from a simple specification test. *Review of Financial Studies* Vol. 1, Issue 1, 41-66.

Merton, Robert C., 1987. A simple model of capital market equilibrium with incomplete information. *Journal of Finance* Vol. XLII, No. 3, 483-510.

Newey, Whitney K., and Kenneth D. West, 1987. A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica* Vol. 55, No. 3, 703-708.

Pastor, Lubos and Robert F. Stambaugh, 2003. Liquidity risk and expected stock return. *Journal of Political Economy* Vol. 111, No. 3, 642-685.

Petersen, Mitchell A. and David Fialkowski, 1994. Posted versus effective spreads: Good prices of bad quotes? *Journal of Financial Economics* 35, 269-292.

Polkovnichenko, Valery, 2005. Household portfolio diversification: A case for rank-dependent preferences. *Review of Financial Studies* Vol. 18, No. 4, 1467-1502.

Table 2.1: Summary statistics of the pooled sample

The monthly sample is based on 9,004 U.S. common stocks listed in NYSE, AMEX, or NASDAQ over the period 1984-2012. Stocks are required to be included in the Russell 3000 index and have at least 100 trades per month. Panel A reports the descriptive statistics of the variables. r is the monthly close-to-close stock return minus risk-free rate. r^{MP} is the monthly quote-midpoint stock return minus risk-free rate. σ_s^2 is the Hasbrouck's (1993) pricing error variance and σ_p^2 is the variance of log transaction prices. $VR(5)$ is weekly-to-daily variance ratio based on daily quote-midpoint return, estimated by the approach of Lo and MacKinlay (1988) using $q = 5$. *Beta*, *MV*, and *B/M* are the stock beta, market value, and book-to-market ratio measured by the approach of Fama and French (1992). *TO* is the monthly turnover ratio, measured by monthly volume scaled by total share outstanding. *IVOL* is the monthly idiosyncratic volatility measured by the approach of Ang et al (2006). *RIVOL* is monthly relative idiosyncratic volatility, measured as the fraction of idiosyncratic daily return volatility to total return volatility over the month. *RES* is monthly trade-weighted averages of relative effective spread. *RQS* is monthly time-weighted average of relative quoted spread. *Amihud* is the Amihud (2002) price impact measure estimated by the approach of Acharya and Pedersen (2005). *Liu* is the Liu (2006) 12-month-no-trade-day measure of illiquidity. *IO* is the quarterly institutional ownership, estimated as the fractions of shares owned by any institutional investors that file the quarterly 13-f reports. *PD* is the Hou and Moskowitz's (2005) measure of price delay (*D1*). *PIN* is Easley, Hvidkjaer, and O'Hara's (2002) annually measure of probability of informed trading. *ANLY* is the number of analysts that report earnings forecasts of a stock.

Variables	Mean	Std dev.	P25	P50	P75	Skew	*N*
Panel A: Descriptive statistics							
Stock Returns							
r	0.83%	14.70%	-6.39%	0.35%	7.27%	0.75	837,779
r^{MP}	0.81%	14.66%	-6.41%	0.32%	7.04%	0.74	836,786
Measures of Price Inefficiency							
σ_s^2/σ_p^2	1.29%	4.30%	0.01%	0.08%	0.56%	6.67	837,779
σ_s^2 ($\times 10^2$)	0.21	1.03	0.00	0.01	0.08	12.48	837,779
$\|VR(5)\|$	0.60	0.74	0.16	0.35	0.70	2.78	786,113
Control Variables							
Beta	1.32	0.41	1.02	1.30	1.61	0.33	730,261
MV ($ billion)	3.43	14.63	0.24	0.59	1.92	14.18	822,050
B/M	64.69%	75.99%	28.83%	50.03%	80.74%	19.72	712,397
TO (/m)	15.19%	17.44%	4.73%	9.55%	19.56%	2.91	837,779
IVOL ($\times 10^2$)	10.61	7.97	5.53	8.47	13.37	2.73	837,774
RIVOL	67.50%	20.87%	53.47%	71.11%	84.58%	-0.58	837,772
EIV	1.78%	3.73%	0.40%	0.82%	1.81%	7.58	664,626
RES	0.83%	0.97%	0.29%	0.56%	1.01%	4.33	837,745
RQS	1.69%	1.65%	0.73%	1.19%	1.97%	3.41	837,090
Amihud	0.87	2.49	0.26	0.32	0.60	8.74	837,774
Liu	4.69	24.28	0.00	0.00	0.00	6.33	837,779
IO	46.91%	32.16%	15.91%	48.14%	72.65%	-0.10	837,779
PD	0.70	0.27	0.50	0.78	0.94	-0.71	772,175
PIN	17.18%	6.62%	12.60%	16.65%	21.60%	0.48	799,986
ANLY	7	7	2	5	10	1.74	837,779

Table 2.2: Determinants of price inefficiency

Coefficient estimates for Fama-MacBeth (1973) cross-sectional regressions of inefficiency on stock characteristics. The monthly sample is based on 9,004 U.S. common stocks listed in NYSE, AMEX, or NASDAQ over the period 1984-2012. Stocks are required to be included in the Russell 3000 index and have at least 100 trades per month. The dependent variable is the natural logarithm of either $(\sigma_s^2/\sigma_p^2)_{t-1}$, the Hasbrouck's (1993) pricing error variance scaled by log price variance, or $|VR(5)_{t-1}|$, the absolute value of weekly-to-daily variance ratio based on daily quote-midpoint return over the previous 3 months. *MV* and *B/M* are the stock market value and book-to-market ratio measured by the approach of Fama and French (1992). RES_{t-1} is trade-weighted average relative effective spread in the previous month. $IVOL_{t-1}$ is idiosyncratic volatility over the previous month. TO_{t-1} is the average monthly turnover ratios over the previous 12 months. IO_{t-1} is institutional ownership at the end of the previous calendar quarter. $ANLY_{t-1}$ is the number of analysts that report earnings forecasts of a stock over the previous two quarters. PRC_{t-1} is the lagged stock price. DV_{t-1} is the dependent variable in the previous month. Ln is the natural logarithm. The *t*-statistics are based on the time-series variation in the regression coefficients over the 347 months and are adjusted by the Newey and West (1987) standard errors.

Dependent Variable	Ln[$(\sigma_s^2/\sigma_p^2)_t$]		Ln($\|VR(5)_t\|$)	
	[1]	[2]	[3]	[4]
Intercept	0.09	-0.09	-0.23	-0.23
t-statistic	0.18	-0.36	-1.86	-1.87
Ln[*MV*]	-0.23	-0.10	-0.08	-0.07
t-statistic	-9.72	-9.46	-8.41	-8.37
Ln[*B/M*]	0.08	0.07	0.04	0.04
t-statistic	4.64	5.56	4.66	4.61
Ln[RES_{t-1}]	0.24	0.18	-0.07	-0.07
t-statistic	3.69	5.00	-5.17	-5.10
Ln[TO_{t-1}]	-0.53	-0.23	-0.02	-0.02
t-statistic	-29.09	-14.85	-3.81	-3.84
Ln[$ANLY_{t-1}$]	-0.11	-0.10	-0.01	-0.01
t-statistic	-10.67	-14.44	-2.86	-2.76
IO_{t-1}	-0.51	-0.33	-0.04	-0.04
t-statistic	-21.93	-27.77	-2.93	-2.96
Ln[PRC_{t-1}]	-0.57	-0.35	0.07	0.07
t-statistic	-11.04	-9.40	4.79	4.79
DV_{t-1}		0.41		0.01
t-statistic		29.68		7.36
Adj. R^2	0.49	0.57	0.02	0.02

Table 2.3: Relation between idiosyncratic volatility and lagged price inefficiency

Coefficient estimates for time-series and Fama-MacBeth (1973) cross-sectional regressions of absolute or relative idiosyncratic volatility on price inefficiency and stock characteristics. The monthly sample is based on 9,004 U.S. common stocks listed in NYSE, AMEX, or NASDAQ over the period 1984-2012. Stocks are required to be included in the Russell 3000 index and have at least 100 trades per month. The dependent variable is either Ln($IVOL$), the natural logarithm of the monthly idiosyncratic volatility measured by the approach of Ang et al (2006), or $RIVOL$, the fraction of idiosyncratic volatility to total return volatility. $(\sigma_s^2/\sigma_p^2)_{t-1}$ is the normalized variant of Hasbrouck's (1993) pricing error variance of the previous month. $VR(5)_{t-1}$ is weekly-to-daily variance ratio based on daily quote-midpoint return over the previous 3 months. DV_{t-1} is the dependent variable in the previous month. $\sigma_{p,t-1}$ is the transaction return volatility over the previous months. *MV* and *B/M* are the stock market value and book-to-market ratio measured by the approach of Fama and French (1992). RES_{t-1} is trade-weighted average relative effective spread in the previous month. TO_{t-1} is the average monthly turnover ratios over the previous 12 months. IO_{t-1} is institutional ownership at the end of the previous calendar quarter. $ANLY_{t-1}$ is the number of analysts that report earnings forecasts of a stock over the previous two quarters. Ln is the natural logarithm. The *t*-statistics of the Fama-MacBeth (1973) cross-sectional regressions are based on the time-series variation in the regression coefficients over the 347 months and are adjusted by the Newey and West (1987) standard errors. To be included in the time-series regression, a stock is required to have at least 50 monthly observations.

	Cross-sectional regressions				Time-series regressions			
Dependent Variable	Ln[$IVOL_t$]		$RIVOL_t$		Ln[$IVOL_t$]		$RIVOL_t$	
Intercept	0.37	0.72	109.24	114.88	-0.20	0.27	135.84	149.21
t-statistic	5.65	15.66	36.50	36.81	-1.78	2.25	28.24	29.26
Ln[$(\sigma_s^2/\sigma_p^2)_{t-1}$]	0.06		1.65		0.12		3.68	
t-statistic	18.11		6.93		14.99		10.26	
Ln[$\lvert VR(5)_{t-1}\rvert$]		0.01		0.07		0.01		0.00
t-statistic		12.53		2.27		16.07		-0.07
DV_{t-1}	0.29	0.32	0.17	0.17	0.07	0.09	0.08	0.09
t-statistic	42.16	33.22	17.60	17.54	19.41	22.59	38.75	38.91
Ln[$\sigma_{p,t-1}$]	0.15	0.08			0.15	0.09		
t-statistic	27.25	15.81			55.66	49.18		
Ln[MV]	-0.03	-0.04	-2.28	-2.43	-0.03	-0.05	-3.83	-4.40
t-statistic	-7.94	-12.00	-15.71	-15.68	-5.90	-8.75	-15.70	-16.97
Ln[B/M]	-0.05	-0.05	-0.50	-0.45	-0.08	-0.09	-1.81	-2.06
t-statistic	-13.65	-13.07	-4.16	-3.55	-13.68	-14.89	-7.70	-8.39
Ln[RES_{t-1}]	0.15	0.19	1.51	1.88	0.13	0.15	-0.54	-0.35
t-statistic	24.38	23.88	10.64	10.80	47.06	58.23	-5.13	-3.33
Ln[TO_{t-1}]	0.16	0.15	0.09	-0.25	0.17	0.16	-1.53	-2.12
t-statistic	32.10	27.18	0.30	-0.88	27.22	25.02	-7.05	-9.23
Ln[$ANLY_{t-1}$]	0.01	0.01	0.15	0.18	-0.02	-0.03	-0.56	-0.67
t-statistic	6.48	6.39	1.56	1.84	-6.67	-6.72	-3.86	-4.37
IO_{t-1}	-0.02	-0.05	-0.80	-1.18	0.02	-0.06	-10.48	-11.88
t-statistic	-2.96	-6.45	-3.02	-4.05	0.84	-2.62	-11.19	-12.16
Adj. R^2	0.55	0.54	0.17	0.17	0.33	0.33	0.18	0.17

Table 2.4: Fama-MacBeth regressions of stock return on price inefficiency

Coefficient estimates for Fama-MacBeth (1973) cross-sectional regressions of stock return on price inefficiency and stock characteristics. The monthly sample is based on 9,004 U.S. common stocks listed in NYSE, AMEX, or NASDAQ over the period 1984-2012. Stocks are required to be included in the Russell 3000 index and have at least 100 trades per month. The dependent variable is monthly stock return in excess of risk-free rate. *Beta*, *MV*, and *B/M* are the beta, market value, and book-to-market ratio measured by the approach of Fama and French (1992). $r_{-2,-12}$ is the compound excess return from month $t-12$ to $t-2$. $(\sigma_s^2/\sigma_p^2)_{t-1}$ is Hasbrouck's (1993) pricing error variance in the previous month scaled by the corresponding return variance. $VR(5)_{t-1}$ is weekly-to-daily variance ratio based on daily quote-midpoint return over the previous 3 months. $IVOL_{t-1}$ is idiosyncratic volatility over the previous month. EIV_t is the expected idiosyncratic risk of the concurrent month estimated by the EGARCH approach of Fu (2009). TO_{t-1} is the average monthly turnover ratios over the previous 12 months. RES_{t-1} is trade-weighted average relative effective spread in the previous month. $Amihud_{t-1}$ is the Amihud (2002) price impact measure of the previous month. *Liu* is the Liu (2006) no-trade-day measure over the previous 12 months. IO_{t-1} is institutional ownership at the end of the previous calendar quarter. Ln is the natural logarithm. The *t*-statistics are based on the time-series variation in the regression coefficients over the 347 months and are adjusted by the Newey and West (1987) standard errors.

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Intercept	2.91	1.41	2.34	3.71	3.38	2.47	4.13	3.01	3.99
t-statistic	3.03	3.53	2.33	3.15	2.89	2.37	3.43	3.15	3.37
Beta	0.03	-0.02	0.07	-0.11	-0.38	-0.09	-0.05	0.04	-0.10
t-statistic	0.10	-0.07	0.27	-0.57	-2.00	-0.49	-0.27	0.15	-0.55
Ln[*MV*]	-0.10		-0.06	-0.08	0.01	-0.04	-0.05	-0.11	-0.10
t-statistic	-2.30		-1.13	-2.03	0.28	-0.92	-1.33	-2.34	-2.48
Ln[*B/M*]	0.15		0.13	0.15	0.23	0.15	0.16	0.13	0.14
t-statistic	1.62		1.46	1.79	2.55	1.85	1.93	1.38	1.69
$r_{-2,-12}$	0.55		0.61	0.54	0.54	0.54	0.41	0.60	0.55
t-statistic	2.33		2.58	2.63	2.69	2.64	1.98	2.50	2.76
Ln[$(\sigma_s^2/\sigma_p^2)_{t-1}$]		0.07	0.08	0.10	0.17	0.08	0.10		
t-statistic		2.18	2.67	3.07	5.90	2.90	3.09		
Ln[$\lvert VR(5)_{t-1}\rvert$]								0.07	0.06
t-statistic								3.28	3.08
Ln[$IVOL_{t-1}$]				-0.58		-0.65	-0.49		-0.67
t-statistic				-5.42		-6.25	-4.06		-6.81
Ln[EIV_t]					0.45				
t-statistic					5.85				
Ln[TO_t]				0.45	0.17	0.57	0.66		0.43
t-statistic				3.85	1.54	4.24	5.67		3.84
Ln[RES_{t-1}]				0.19	-0.26				0.32
t-statistic				2.20	-2.84				3.33
Ln[$Amihud_{t-1}$]						0.60			
t-statistic						4.01			
Ln[Liu_{t-1}]							0.15		
t-statistic							8.83		
IO_{t-1}				0.25	0.33	0.24	0.22		0.17
t-statistic				1.89	2.41	1.84	1.67		1.26
Adj. R^2	0.05	0.03	0.05	0.07	0.07	0.07	0.07	0.05	0.07

Table 2.5: Returns and characteristics of portfolios formed on inefficiency measures

The monthly sample is based on 9,004 U.S. common stocks listed in NYSE, AMEX, or NASDAQ over the period 1984-2012. Stocks are required to be included in the Russell 3000 index and have at least 100 trades per month. At the beginning of each month, stocks are sorted into value-weighted quintile portfolios by price inefficiency of the previous month. The time-series average of the excess returns (r), *FF3*-adjusted returns (α_{FF3}), and *FF3*-and-momentum-adjusted returns (α_{FF4}) are reported for each quintile portfolio as well as a hypothetical zero-investment portfolio that is long in the High quintile and short in the Low quintile (High – Low). Panel A reports the returns of quintile portfolios sorted by $(\sigma_s^2/\sigma_p^2)_{t-1}$, the normalized variant of Hasbrouck's (1993) pricing error variance. Panel B presents the value-weighted average of stock characteristics of each portfolio. *MV* and *B/M* are the stock market value and book-to-market ratio measured by the approach of Fama and French (1992). $r_{-2,-12}$ is the compound excess return from month $t-12$ to $t-2$. $IVOL_t$ and $IVOL_{t-1}$ are idiosyncratic volatility in the concurrent and previous months, respectively. RES_{t-1} is trade-weighted average relative effective spread in the previous month. TO_{t-1} is turnover ratios of the previous month. $ANLY_{t-1}$ is the number of analysts that report earnings forecasts of a stock over the previous two quarters. IO_{t-1} is institutional ownership at the end of the previous calendar quarter. Panel C reports the returns of quintile portfolios sorted by $|VR(5)_{t-1}|$, where $VR(5)_{t-1}$ is weekly-to-daily variance ratio based on daily quote-midpoint return over the previous 3 months. The *t*-statistics are based on the time-series variation in portfolio returns over the 347 months and are adjusted by the Newey and West (1987) standard errors.

	Low	2	3	4	High	High – Low
Panel A: Returns of quintile portfolios sorted by $(\sigma_s^2/\sigma_p^2)_{t-1}$.						
Portfolio returns are based on closing prices						
r	0.53%	0.55%	0.70%	0.88%	0.93%	0.40%
t-statistic	1.90	2.19	2.88	3.54	3.44	2.12
α_{CAPM}	-0.09%	-0.04%	0.14%	0.34%	0.42%	0.51%
t-statistic	-1.93	-0.73	1.62	2.70	2.55	2.80
α_{FF3}	-0.09%	-0.03%	0.09%	0.25%	0.25%	0.35%
t-statistic	-1.80	-0.60	1.15	2.41	2.14	2.45
α_{FF4}	-0.11%	-0.02%	0.12%	0.32%	0.38%	0.50%
t-statistic	-2.16	-0.38	1.58	3.01	3.60	3.95
$\alpha_{FF4,EW}$	-0.11%	0.01%	0.11%	0.31%	0.59%	0.70%
t-statistic	-1.62	0.30	1.86	4.00	5.06	4.83
Portfolio returns are based on quote midpoints						
r^{MP}	0.50%	0.54%	0.71%	0.87%	0.93%	0.43%
t-statistic	1.81	2.17	2.93	3.52	3.48	2.27
α_{CAPM}	-0.11%	-0.04%	0.15%	0.34%	0.43%	0.54%
t-statistic	-2.11	-0.75	1.82	2.70	2.62	2.96
α_{FF3}	-0.11%	-0.04%	0.10%	0.25%	0.26%	0.38%
t-statistic	-2.05	-0.65	1.33	2.41	2.23	2.67
α_{FF4}	-0.12%	-0.02%	0.14%	0.32%	0.40%	0.52%
t-statistic	-2.18	-0.30	1.93	3.08	3.83	4.17
$\alpha_{FF4,EW}$	-0.13%	0.01%	0.10%	0.31%	0.58%	0.70%
t-statistic	-1.82	0.23	1.78	4.03	5.01	4.88

Table 2.5: Continued

	Low	2	3	4	High	High – Low
Panel B: Average stock characteristics of quintile portfolios sorted by $MISPRC_{t-1}$.						
$MISPRC_{t-1}$	0.05%	0.16%	0.42%	1.12%	6.16%	6.11%
$\sigma_{s,t-1}$	0.08%	0.14%	0.22%	0.33%	0.61%	0.53%
$\sigma_{p,t-1}$	6.22%	5.09%	4.83%	4.48%	3.68%	-2.53%
MV (\$ *billion*)	6.59	4.62	2.67	1.29	0.51	-6.08
B/M	56.12%	61.39%	67.38%	74.13%	84.35%	28.24%
$r_{-2,-12}$	21.76%	17.01%	12.44%	5.77%	-2.36%	-24.12%
$IVOL_t$ ($\times 10^2$)	8.56	9.38	10.17	11.00	12.26	3.70
$IVOL_{t-1}$ ($\times 10^2$)	9.61	9.81	10.29	10.65	11.02	1.41
RES_t	0.46%	0.61%	0.79%	1.03%	1.61%	1.15%
TO_t ($/m$)	18.38%	16.53%	14.37%	11.47%	7.55%	-10.83%
$ANLY_t$	9.02	7.14	5.41	3.88	2.30	-6.72
IO_t	55.08%	50.73%	45.91%	39.33%	29.20%	-25.88%
Panel C: Returns of quintile portfolios sorted by $\|VR(5)_{t-1}\|$.						
Portfolio returns are based on closing prices						
r	0.40%	0.54%	0.67%	0.57%	0.75%	0.36%
t-statistic	1.34	2.09	2.64	2.23	2.77	2.16
α_{CAPM}	-0.21%	-0.04%	0.09%	-0.02%	0.12%	0.33%
t-statistic	-1.62	-0.61	1.15	-0.27	1.31	1.87
α_{FF3}	-0.22%	-0.05%	0.08%	-0.01%	0.15%	0.37%
t-statistic	-1.84	-0.84	0.98	-0.18	1.52	2.13
α_{FF4}	-0.21%	-0.05%	0.06%	-0.01%	0.15%	0.37%
t-statistic	-1.70	-0.62	0.88	-0.14	1.64	2.09
$\alpha_{FF4,EW}$	-0.13%	0.03%	0.14%	0.28%	0.49%	0.62%
t-statistic	-1.15	0.46	2.64	3.79	4.53	3.83
Portfolio returns are based on quote midpoints						
r	0.37%	0.54%	0.65%	0.56%	0.72%	0.35%
t-statistic	1.26	2.11	2.54	2.20	2.64	2.13
α_{CAPM}	-0.22%	-0.03%	0.08%	-0.02%	0.10%	0.31%
t-statistic	-1.68	-0.48	0.96	-0.33	1.06	1.84
α_{FF3}	-0.24%	-0.05%	0.06%	-0.02%	0.12%	0.36%
t-statistic	-1.95	-0.75	0.76	-0.28	1.24	2.12
α_{FF4}	-0.22%	-0.03%	0.06%	-0.01%	0.14%	0.36%
t-statistic	-1.76	-0.43	0.77	-0.10	1.45	2.08
$\alpha_{FF4,EW}$	-0.13%	0.03%	0.12%	0.27%	0.47%	0.60%
t-statistic	-1.15	0.50	2.15	3.76	4.42	3.81

Table 2.6: Portfolio based on independent double sort by σ_s^2/σ_p^2 and IO

Time-series averages of the four-factor-adjusted returns (α_{FF4}) of value-weighted portfolios sorted by σ_s^2/σ_p^2, conditional on institutional ownership (IO). The monthly sample is based on 9,004 U.S. common stocks listed in NYSE, AMEX, or NASDAQ over the period 1984-2012. Stocks are required to be included in the Russell 3000 index and have at least 100 trades per month. Stocks are sorted into 20 portfolios by $(\sigma_s^2/\sigma_p^2)_{t-1}$ and IO_{t-1} independently at the beginning of each month. $(\sigma_s^2/\sigma_p^2)_{t-1}$ is Hasbrouck's (1993) pricing error variance in the previous month scaled by the corresponding return variance. IO_{t-1} is institutional ownership at the end of the previous calendar quarter. The time-series average of the value-weighted α_{FF4} are reported for each of the 20 portfolios as well as 4 hypothetical zero-investment portfolios that is long the High σ_s^2/σ_p^2 stocks and short in the Low σ_s^2/σ_p^2 stocks (High – Low). Panel A, B, and C are based on the whole stock sample, NYSE stocks, and Russell 2000 stocks, respectively. The t-statistics are based on the time-series variation in portfolio returns over the 347 months and are adjusted by the Newey and West (1987) standard errors.

	σ_s^2/σ_p^2 Low	2	3	4	σ_s^2/σ_p^2 High	High – Low
Panel A: The whole sample						
IO Low	-0.27%	-0.20%	-0.02%	0.16%	0.34%	0.61%
t-statistic	-1.78	-1.37	-0.11	1.11	2.20	2.90
2	-0.24%	-0.10%	0.04%	0.20%	0.55%	0.79%
t-statistic	-1.27	-0.46	0.27	1.37	4.47	3.57
3	-0.10%	0.06%	0.20%	0.50%	0.57%	0.67%
t-statistic	-1.20	0.77	1.93	3.29	3.71	3.51
IO High	-0.09%	0.04%	0.11%	0.13%	0.12%	0.21%
t-statistic	-0.87	0.46	1.00	0.88	0.70	1.22
Panel B: NYSE stocks only						
IO Low	-0.23%	-0.10%	0.01%	0.23%	0.33%	0.57%
t-statistic	-1.36	-0.61	0.07	1.50	1.67	2.33
2	-0.42%	-0.04%	0.20%	0.19%	0.63%	1.05%
t-statistic	-2.56	-0.21	1.13	0.92	2.79	3.59
3	-0.13%	0.10%	0.16%	0.51%	0.65%	0.78%
t-statistic	-1.50	1.04	1.32	2.85	3.13	3.55
IO High	-0.10%	0.00%	0.10%	0.11%	0.03%	0.13%
t-statistic	-0.87	0.04	0.71	0.63	0.10	0.50
Panel C: Russell 2000 stocks only						
IO Low	-1.74%	-1.27%	-0.53%	-0.55%	-0.12%	1.65%
t-statistic	-6.25	-6.38	-2.58	-3.77	-0.74	4.97
2	-0.76%	-0.64%	-0.46%	0.23%	0.38%	1.08%
t-statistic	-3.43	-3.76	-4.11	1.01	3.07	4.01
3	-0.28%	-0.48%	-0.19%	0.00%	0.38%	0.65%
t-statistic	-1.29	-3.28	-1.60	0.01	2.58	2.31
IO High	-0.55%	-0.55%	-0.22%	-0.12%	0.02%	0.60%
t-statistic	-2.62	-3.88	-1.68	-0.80	0.09	2.43

Table 2.7: Robustness tests

The monthly sample is based on 9,004 U.S. common stocks listed in NYSE, AMEX, or NASDAQ over the period 1984-2012. Stocks are required to be included in the Russell 3000 index and have at least 100 trades per month. At the beginning of each month, stocks are sorted into quintile portfolios by $(\sigma_s^2/\sigma_p^2)_{t-1}$, the normalized variant of Hasbrouck's (1993) pricing error variance. The time-series averages of the value-weighted *FF3*-and-momentum-adjusted returns (α_{FF4}) are reported. *MV* is market value of a stock measured by the approach of Fama and French (1992). *RES* is monthly trade-weighted averages of relative effective spread. *RQS* is monthly time-weighted average of relative quoted spread. *Amihud* is the Amihud (2002) price impact measure estimated by the approach of Acharya and Pedersen (2005). *Liu* is the Liu (2006) 12-month-no-trade-day measure of illiquidity. The *P-S* traded factor is the Pastor-Stambaugh (2003) traded liquidity risk factor. The Liu traded factor is the Liu (2006) liquidity factor based on 12-month no-trade-day illiquidity measure. *IVOL* is the monthly idiosyncratic volatility measured by the approach of Ang et al (2006). *PD* is the Hou and Moskowitz's (2005) measure of price delay (*D1*). *PIN* is Easley, Hvidkjaer, and O'Hara's (2002) annually measure of probability of informed trading. Panel A controls for stock size by testing on sub-samples of constituents of certain market indexes and on size quartiles. Panel B controls for illiquidity and transaction cost by testing on sub-samples of stocks based on various illiquidity and transaction cost measures and by adding an illiquidity factor into the four-factor regression. Panel C controls for IVOL of the previous month. Panel D controls for potential price-delay effect by testing on low and high *PD* quartiles. Panel E controls for potential information risk by testing on low and high *PIN* quartiles. Panel F tests many alternative specifications: *i*) use the original Hasbrouck's (1993) pricing error variance instead of its normalized variant, *ii*) use NYSE breakpoints of σ_s^2/σ_p^2 and market share breakpoints, *iii*) skips 1 month between the ranking period and the holding period, *iv*) drop stocks with the greatest price inefficiency, *v*) test on sub-periods before and after the decimalization in 2001, or drop the financial crisis years, *vi*) drops all January observations to control for potential January effect. The *t*-statistics are based on the time-series variation in portfolio returns over the 347 months and are adjusted by the Newey and West (1987) standard errors.

	Low	2	3	4	High	High – Low
Panel A: Control for size						
MV Low	-0.09%	0.01%	0.09%	0.24%	0.32%	0.41%
t-statistic	-0.79	0.13	0.90	2.03	1.93	1.90
2	-0.12%	-0.23%	-0.09%	0.13%	0.31%	0.43%
t-statistic	-1.15	-3.05	-1.00	1.53	3.38	2.62
3	-0.13%	-0.05%	0.02%	0.17%	0.23%	0.36%
t-statistic	-1.20	-0.55	0.28	2.30	2.27	2.57
MV High	-0.09%	-0.10%	0.06%	0.00%	0.28%	0.37%
t-statistic	-1.02	-1.30	0.96	-0.05	2.55	2.52
S&P 500 Only	-0.09%	-0.02%	0.06%	0.10%	0.22%	0.31%
t-statistic	-1.15	-0.30	1.03	1.16	2.02	2.22
Russell 1000 Only	-0.08%	-0.05%	0.03%	0.13%	0.28%	0.36%
t-statistic	-1.06	-0.62	0.50	1.48	2.81	2.63
Russell 2000 Only	-0.71%	-0.43%	-0.14%	0.02%	0.13%	0.85%
t-statistic	-6.51	-4.71	-1.50	0.17	0.73	3.47
NYSE Only	-0.18%	0.01%	0.07%	0.11%	0.49%	0.67%
t-statistic	-2.70	0.16	0.74	0.99	4.05	4.75
w/ Non-Russell 3000 Stocks	-0.08%	0.00%	0.30%	0.29%	0.44%	0.52%
t-statistic	-2.01	-0.06	2.85	2.58	3.57	3.97

Table 2.7: Continued

	Low	2	3	4	High	High – Low
Panel B: Control for lagged illiquidity and transaction cost						
RES Low Quartile	-0.13%	-0.06%	0.20%	-0.08%	0.18%	0.31%
t-statistic	-1.63	-0.81	2.65	-0.90	1.83	2.39
RES High Quartile	-0.18%	0.03%	0.07%	0.28%	0.45%	0.63%
t-statistic	-1.14	0.17	0.41	1.42	2.17	2.58
Amihud Low Quartile	-0.12%	-0.05%	0.10%	-0.03%	0.27%	0.38%
t-statistic	-1.40	-0.79	1.68	-0.38	2.61	2.68
Amihud High Quartile	-0.17%	-0.13%	0.02%	0.18%	0.51%	0.68%
t-statistic	-1.47	-1.44	0.16	1.29	2.92	3.23
Price ≥ $5 Only	-0.13%	-0.03%	-0.03%	0.11%	0.35%	0.49%
t-statistic	-2.01	-0.49	-0.43	1.39	3.12	3.49
Large/Liquid	-0.10%	-0.03%	-0.01%	-0.01%	0.18%	0.27%
t-statistic	-1.31	-0.35	-0.15	-0.17	2.08	2.32
Large/High Price	-0.14%	0.06%	-0.13%	0.09%	0.25%	0.38%
t-statistic	-2.20	1.01	-1.68	1.08	2.41	3.14
Liquid/High Price	-0.11%	0.00%	0.01%	0.31%	0.31%	0.43%
t-statistic	-2.09	0.04	0.09	2.90	3.05	3.50
Control for *P-S Traded-Factor*	-0.14%	-0.02%	0.14%	0.30%	0.39%	0.53%
t-statistic	-2.47	-0.28	1.71	2.81	3.64	4.13
Control for *Liu Traded-Factor*	-0.09%	-0.02%	0.13%	0.36%	0.34%	0.43%
t-statistic	-1.48	-0.39	1.56	3.17	2.88	3.03
Panel C: Control for lagged idiosyncratic volatility						
IVOL Low	-0.06%	0.11%	0.17%	0.33%	0.38%	0.44%
t-statistic	-0.74	1.06	1.46	2.76	3.14	3.32
2	-0.02%	-0.09%	-0.12%	0.05%	0.37%	0.40%
t-statistic	-0.29	-0.87	-0.84	0.33	2.51	2.42
3	0.00%	0.15%	0.18%	0.33%	0.51%	0.51%
t-statistic	0.01	0.99	1.01	2.08	3.70	2.13
IVOL High	-0.35%	-0.39%	-0.21%	-0.42%	-0.12%	0.23%
t-statistic	-1.52	-1.72	-0.94	-2.21	-0.57	0.76
Panel D: Control for price delay *(PD)*						
PD Low Quartile	0.02%	-0.02%	0.00%	0.56%	0.42%	0.40%
t-statistic	0.23	-0.13	0.03	3.36	3.26	2.57
PD High Quartile	-0.24%	0.02%	0.11%	0.32%	0.26%	0.50%
t-statistic	-2.92	0.16	0.77	1.85	1.97	3.27
Panel E: Control for *PIN*						
PIN Low Quartile	-0.13%	-0.02%	-0.06%	0.12%	0.35%	0.48%
t-statistic	-1.46	-0.26	-0.85	1.18	2.43	2.76
PIN High Quartile	-0.19%	-0.05%	0.14%	0.09%	0.52%	0.71%
t-statistic	-1.11	-0.40	1.15	0.82	3.47	2.59

Table 2.7: Continued

	Low	2	3	4	High	High – Low
Panel F: Alternative specifications						
Form portfolios by the original Hasbrouck's pricing error variance (σ_s^2)						
α_{FF4} based on closing-price return	-0.01%	-0.03%	0.16%	0.09%	0.28%	0.29%
t-statistic	-0.27	-0.66	1.93	0.88	2.13	2.06
α_{FF4} based on quote-midpoint return	-0.01%	-0.03%	0.18%	0.10%	0.29%	0.31%
t-statistic	-0.34	-0.58	2.09	0.94	2.28	2.20
Alternative breakpoints						
NYSE σ_s^2/σ_p^2 breakpoint	-0.12%	0.00%	0.07%	0.15%	0.45%	0.58%
t-statistic	-2.18	-0.06	0.95	1.52	4.26	4.46
Market share breakpoint	-0.10%	-0.13%	0.13%	-0.01%	0.23%	0.33%
t-statistic	-1.35	-2.04	2.15	-0.12	2.84	2.82
Skip 1 month between the ranking period and the holding period						
α_{FF4} based on closing-price return	0.02%	0.04%	-0.12%	0.17%	0.27%	0.25%
t-statistic	0.43	0.83	-1.80	1.65	2.76	2.26
α_{FF4} based on quote-midpoint return	-0.07%	0.03%	-0.17%	0.15%	0.40%	0.47%
t-statistic	-1.14	0.48	-2.37	1.44	2.70	2.92
Drop stocks with the greatest price inefficiency						
Drop top 10% by σ_s^2/σ_p^2	-0.12%	-0.01%	0.03%	0.38%	0.37%	0.49%
t-statistic	-2.19	-0.17	0.35	3.49	3.39	3.76
Drop top 20% by σ_s^2/σ_p^2	-0.12%	-0.01%	-0.04%	0.23%	0.27%	0.39%
t-statistic	-2.08	-0.17	-0.52	2.77	2.37	2.89
Subperiods of pre- and post-decimalization						
1984-2000	-0.04%	-0.06%	0.02%	0.19%	0.38%	0.43%
t-statistic	-0.79	-0.75	0.22	1.43	3.00	2.94
2001-2012	-0.20%	0.04%	0.21%	0.35%	0.33%	0.53%
t-statistic	-2.42	0.55	2.20	2.71	2.09	2.71
Drop 2008-2009	-0.10%	0.00%	0.10%	0.31%	0.43%	0.53%
t-statistic	-2.04	-0.08	1.29	2.81	3.94	4.20
Drop Januarys						
α_{FF4}	-0.11%	0.01%	0.11%	0.32%	0.35%	0.46%
t-statistic	-1.88	0.16	1.58	3.33	3.55	3.86

Table 2.8: Portfolio returns based on simulated pricing error

Stochastic pricing errors are simulated for a number of stocks over 348 periods. Pricing error of each stock is assumed to follow a lognormal distribution with zero mean and certain volatility σ_z. Market values of the S&P 500 stocks at the end of year 2011 are assigned to the simulated stocks. To simplify the simulation process and to eliminate potential complications, all stocks are assumed to have zero return on their efficient price thus have no change in their intrinsic market values. The volatility of pricing errors is assumed to range from 0.5% to 20%. The time-series average of the equally-weighted and value-weighted portfolios returns based on simulated stock prices are reported. Panel A reports results for portfolios with 500 stocks. Panel B presents results for portfolios with only 100 stocks. The *t*-statistics are based on the time-series variation in portfolio returns over the 348 months and are adjusted by the Newey and West (1987) standard errors.

	Volatility of Pricing Error (σ_z)							
	0.5%	1%	2%	3%	5%	10%	15%	20%
Panel A: The portfolio has 500 stocks								
EW Port. Ret.	0.00%	0.01%	0.04%	0.09%	0.25%	1.00%	2.27%	4.07%
t-statistic	3.54	7.08	14.08	21.08	34.95	66.65	91.75	112.95
VW Port. Ret.	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.02%	0.03%
t-statistic	0.02	0.03	0.06	0.08	0.14	0.28	0.40	0.57
Panel B: The portfolio has 100 stocks								
EW Port. Ret.	0.00%	0.01%	0.04%	0.09%	0.25%	1.00%	2.27%	4.07%
t-statistic	1.57	3.16	6.33	9.43	15.70	30.01	42.54	52.55
VW Port. Ret.	0.00%	0.00%	0.00%	0.00%	0.01%	0.03%	0.08%	0.14%
t-statistic	0.02	0.06	0.12	0.18	0.29	0.58	0.86	1.16

Table 2.9: Portfolio returns net of transaction costs

The monthly sample is based on 9,004 U.S. common stocks listed in NYSE, AMEX, or NASDAQ over the period 1984-2012. Stocks are required to be included in the Russell 3000 index and have at least 100 trades per month. At the beginning of each month, stocks are sorted into portfolios by $(\sigma_s^2/\sigma_p^2)_{t-1}$, the normalized variant of Hasbrouck's (1993) pricing error variance in the previous month. Transaction cost (TC) is measured by either relative effective spread (*RES*) or relative quoted spread (*RQS*). Low TC stocks are defined as the bottom 50% stocks sorted by their transaction costs in the previous month. The time-series average of the value-weighted *FF3*-and-momentum-adjusted net returns (α_{FF4}) are reported. The *t*-statistics are based on the time-series variation in portfolio returns over the 347 months and are adjusted by the Newey and West (1987) standard errors.

	TC measured by effective spread				TC measured by quote spread			
Rebalancing interval (*months*)	1	3	6	12	1	3	6	12
Long in top 10%, short in bottom 10%	0.14%	0.50%	0.02%	0.03%	-0.69%	0.20%	-0.14%	-0.07%
t-statistic	0.58	1.85	0.10	0.18	-2.61	0.72	-0.89	-0.42
Long in top 20%, short in bottom 20%	0.13%	0.41%	0.27%	0.25%	-0.46%	0.19%	0.16%	0.18%
t-statistic	0.68	2.20	1.44	1.50	-2.27	1.00	0.85	1.07
Long in top 30%, short in bottom 30%	0.14%	0.32%	0.17%	0.26%	-0.29%	0.17%	0.10%	0.21%
t-statistic	1.07	2.38	1.34	2.10	-2.02	1.22	0.73	1.71
Long in top 50%, short in bottom 50%	0.16%	0.23%	0.15%	0.17%	-0.14%	0.12%	0.09%	0.14%
t-statistic	1.59	2.53	1.57	1.67	-1.29	1.27	0.99	1.33
S&P 500 stocks	0.21%	0.28%	0.28%	0.22%	-0.19%	0.13%	0.21%	0.18%
t-statistic	1.14	1.65	1.71	1.28	-1.00	0.79	1.29	1.03
Russell 2000 stocks	-0.06%	0.57%	0.71%	0.77%	-1.06%	0.19%	0.51%	0.64%
t-statistic	-0.21	3.20	3.67	3.93	-3.26	1.05	2.57	3.18
Low TC stocks	0.38%	0.34%	0.25%	0.29%	-0.05%	0.23%	0.12%	0.14%
t-statistic	2.03	2.30	1.74	1.87	-0.24	1.38	0.79	0.95

Table 2.10: Contributions of price inefficiency premium to traditional risk factors

The monthly sample is based on all U.S. common stocks listed in NYSE, AMEX, or NASDAQ over the period 1984-2012. The inefficiency premium (*IP*) is measured by the difference between returns of the top quintile of stocks sorted by $(\sigma_s^2/\sigma_p^2)_{t-1}$ and that of the bottom quintile, rebalanced monthly. Panel A presents the value-weighted excess returns (r), market-adjusted returns (α_{CAPM}), and market-and-IP-adjusted returns (α_2) of portfolios classified by market values. Panel B reports those of portfolios classified by book-to-market ratios. Panel C presents those of portfolios sorted by compound excess returns from month $t-12$ to month $t-2$. The *t*-statistics are based on the time-series variation in portfolio returns over the 347 months and are adjusted by the Newey and West (1987) standard errors.

	Big	2	3	4	Small	Small – Big
Panel A: *MV*-sorted quintile portfolios						
Raw and CAPM-adjusted returns, and VW average σ_s^2/σ_p^2						
r	0.67%	0.75%	0.75%	0.89%	1.43%	0.76%
t-statistic	2.74	2.54	2.24	2.29	2.95	2.23
α_{CAPM}	0.11%	0.10%	0.06%	0.19%	0.72%	0.61%
t-statistic	1.83	0.82	0.38	0.83	2.30	1.88
σ_s^2/σ_p^2	0.28%	0.92%	1.80%	3.19%	5.45%	5.17%
t-statistic	6.65	7.22	7.96	10.45	15.55	16.23
Two-factor-adjusted performance						
α_2	0.13%	-0.11%	-0.30%	-0.32%	0.06%	-0.07%
t-statistic	2.09	-1.06	-2.23	-1.86	0.25	-0.25
β_m	0.94	1.22	1.34	1.41	1.48	0.54
t-statistic	67.91	49.33	43.01	35.21	27.43	8.91
β_{IP}	-0.02	0.35	0.60	0.86	1.10	1.12
t-statistic	-1.05	9.90	13.34	14.82	13.99	12.66
Adj R^2	0.94	0.88	0.84	0.78	0.69	0.34

	B/M Low	2	3	4	*B/M* High	High – Low
Panel B: *B/M* portfolios						
Raw and CAPM-adjusted returns, and VW average σ_s^2/σ_p^2						
r	0.40%	0.76%	0.89%	1.07%	1.19%	0.79%
t-statistic	1.10	2.45	3.07	3.58	3.57	3.12
α_{CAPM}	-0.30%	0.11%	0.28%	0.50%	0.59%	0.90%
t-statistic	-1.76	0.80	2.04	2.98	2.91	3.32
σ_s^2/σ_p^2	0.29%	0.36%	0.44%	0.63%	0.88%	0.60%
t-statistic	6.07	6.60	7.93	8.07	8.04	8.75
Two-factor-adjusted performance						
α_2	-0.53%	-0.14%	0.01%	0.22%	0.27%	0.80%
t-statistic	-3.42	-1.15	0.14	2.25	2.18	3.99
β_m	1.32	1.23	1.16	1.09	1.14	-0.18
t-statistic	36.93	45.03	53.43	48.14	40.08	-3.85
β_{IP}	0.38	0.40	0.45	0.47	0.54	0.15
t-statistic	7.37	10.20	14.29	14.22	13.00	2.31
Adj R^2	0.80	0.86	0.89	0.87	0.82	0.08

Table 2.10: Continued

	$r_{-2,-12}$ Low	2	3	4	$r_{-2,-12}$ High	High – Low
Panel C: Momentum portfolios						
Raw and CAPM-adjusted returns, and VW average σ_s^2/σ_p^2						
r	0.13%	0.52%	0.53%	0.71%	0.84%	0.71%
t-statistic	0.26	1.68	2.08	3.14	2.68	1.69
α_{CAPM}	-0.74%	-0.12%	-0.02%	0.19%	0.21%	0.96%
t-statistic	-2.58	-0.76	-0.26	2.63	1.37	2.52
σ_s^2/σ_p^2	1.14%	0.57%	0.46%	0.41%	0.40%	-0.74%
t-statistic	6.96	7.26	7.61	6.74	6.52	-5.84
Two-factor-adjusted performance						
α_2	-0.98%	-0.19%	-0.05%	0.23%	0.21%	1.18%
t-statistic	-3.35	-1.12	-0.53	2.78	1.26	2.93
β_m	1.59	1.13	0.95	0.87	1.09	-0.51
t-statistic	23.79	29.30	44.96	46.35	29.00	-5.47
β_{IP}	0.38	0.11	0.04	-0.06	0.04	-0.35
t-statistic	3.95	1.91	1.29	-2.34	0.67	-2.58
Adj R^2	0.63	0.73	0.87	0.88	0.73	0.08

Table 2.A.1: σ_s^2/σ_p^2 and Profits of Momentum and Reversal Strategies

Value-weighted momentum and reversal portfolios are built on subsamples of stocks sorted by average σ_s^2/σ_p^2, and the time-series averages of the three-factor-adjusted portfolio returns are reported. σ_s^2/σ_p^2 is the Hasbrouck's (1993) pricing error variance scaled by the corresponding return variance. The stock sample is based on 9,004 U.S. common stocks listed in NYSE, AMEX, or NASDAQ over the period 1984-2012. Stocks are required to be included in the Russell 3000 index and have at least 100 trades per month. $r_{-2,-13}$ is the compound excess return from month $t-13$ to $t-2$. $r_{w,-2}$ is the excess return of week $t-2$. In Panel A, at the beginning of each month, stocks are sorted into 25 portfolios by $r_{-2,-12}$ and average σ_s^2/σ_p^2 over the current year independently, and are held for one month. In Panel B, at the beginning of each week, stocks are sorted into 25 portfolios by $r_{w,-2}$ and average σ_s^2/σ_p^2 over the current quarter independently, and are held for one week. The *t*-statistics are based on the time-series variation in portfolio returns and are adjusted by the Newey and West (1987) standard errors.

Panel A: Momentum Strategy

	$r_{-2,-13}$ Low	2	3	4	$r_{-2,-13}$ High	High – Low
σ_s^2/σ_p^2 Low	0.17%	0.06%	0.03%	0.15%	0.45%	0.28%
t-statistic	0.57	0.42	0.33	1.78	2.69	0.68
2	-0.51%	-0.31%	-0.17%	0.01%	-0.05%	0.46%
t-statistic	-1.95	-2.28	-1.26	0.10	-0.30	1.25
3	-1.09%	-0.24%	-0.18%	-0.08%	-0.23%	0.86%
t-statistic	-3.73	-1.42	-1.17	-0.70	-1.48	2.19
4	-1.42%	-0.41%	-0.01%	-0.09%	-0.31%	1.11%
t-statistic	-4.48	-2.26	-0.07	-0.69	-1.69	2.88
σ_s^2/σ_p^2 High	-1.81%	-0.19%	0.02%	0.16%	-0.48%	1.32%
t-statistic	-6.12	-1.25	0.12	1.31	-2.34	3.26

Panel B: Reversal Strategy

	$r_{w,-2}$ High	2	3	4	$r_{w,-2}$ Low	Low – High
σ_s^2/σ_p^2 Low	0.03%	0.00%	0.04%	0.09%	0.02%	0.00%
t-statistic	0.63	-0.13	1.68	3.32	0.47	-0.07
2	-0.06%	-0.12%	0.03%	0.05%	0.00%	0.06%
t-statistic	-1.63	-4.20	1.33	1.69	0.04	0.90
3	-0.16%	-0.08%	0.00%	-0.01%	-0.04%	0.12%
t-statistic	-4.26	-2.88	0.05	-0.19	-0.92	1.92
4	-0.18%	-0.10%	-0.02%	-0.03%	-0.05%	0.13%
t-statistic	-4.90	-3.50	-0.67	-0.91	-0.95	2.18
σ_s^2/σ_p^2 High	-0.31%	-0.05%	-0.03%	-0.11%	-0.16%	0.16%
t-statistic	-7.96	-1.54	-1.16	-3.14	-3.90	3.06

Essay 3: Investor Portfolios When Stocks Are Mispriced

1. Introduction

With the growth in trading of stocks, Charles Dow and Edward Jones introduced a simple way in the 1880s to keep track of the market: just sum the prices of key securities. Academic research and practitioner wisdom evolved to create value-weighted indices such as the S&P 90 in 1923. Portfolio theory introduced in the 1950s showed that diversification of risk by holding a portfolio of different kinds of securities generated a superior risk-return tradeoff than holding a small number of related securities, and confirmed the dominance of value-weighted indices. The first mutual funds as a vehicle for investment were introduced in Boston in 1924. As the number of mutual funds increased, investors began using indices to judge the performance of and compare mutual funds. Since much academic evidence had been presented in the 1950s and 1960s to show that mutual funds can't beat indices, a natural progression was the introduction of index funds. The first index fund was introduced by Wells Fargo in 1973. Currently, about 8% of invested capital is passively indexed to value-weighted equity indices.

Indexed funds, both mutual funds and exchange traded funds, have several desirable characteristics. First, the management fees can be quite low because the managers are not expected to exhibit any stock-picking skills. In large cap funds (like the Vanguard 500 Index Admiral Shares [VFIAX], Fidelity Spartan 500 Index [FUSEX]) these fees do not usually exceed 10 basis points, or 0.10% a year. The second desirable characteristic of indexed mutual funds is low turnover of portfolio holdings. Turnover is critical to performance of mutual funds. Typical mutual funds have an average turnover of 100% a year. An average mutual fund loses 0.50% to 1.00% of its annual returns to trading costs. On the other hand, the annual turnover rates of large index funds average less than 10% a year, possibly less than 5% a year. Third, the tax liability of passive investors depends on dividends and capital gains realized by the fund during the year. Lower turnover means that indexed funds do not trade as much as actively managed funds, implying realization of significantly smaller gains. As a result, smaller capital gains are distributed to long-term investors, which translate into a lower current tax liability for investors.

The historical perspective illustrates that investors began investing in value-weighted portfolios, not by design but by default: the indices were available. Over the last few decades, index providers have modified the indices to be investor friendly encouraging further use of indices as investment vehicles.

The question we seek to answer in this paper is whether value-weighted indices are optimal for individual investors. We are not the first to raise this question. Several studies in recent years have documented outperformance of equally-weighted (EW) portfolios compared with value-weighted (VW) portfolios both in raw returns and risk-adjusted returns. Hsu (2006) suggests that VW portfolios are sub-optimal due to potential errors in stock prices. As the weight of a stock in a VW portfolio depends on its current market capitalization, it will be higher in overvalued stocks that are likely to have lower future returns, and lower in undervalued stocks that are likely to have higher future returns, resulting in suboptimal performance. Arnott (2006) and Siegel (2006), among several others, suggests the same explanation. However, Perold (2007) suggests that Hsu's (2006) model is flawed and that the negative abnormal returns of VW portfolios suggested by Hsu (2006) does not exist under a more realistic model setting.

In this paper, we re-examine the outperformance of the EW indexes and aim to provide a more reasonable explanation for this puzzling phenomenon. Specifically, we show that such an outperformance is probably due to the "mispricing return premium" proposed by Blume and Stambaugh (1983) and Brennan and Wang (2010). Due to Jensen's inequality, random errors in stock price could result in a higher rate of holding period return for individual stocks. As long as stock prices are not fully efficient, returns of EW portfolios, which rely on regularly rebalancing, will also be biased upward. On the other hand, VW portfolios represent a buy-and-hold strategy thus are largely unaffected by such a return bias. Therefore, the observed outperformance of EW portfolios may be a manifestation of the mispricing return premium, but represents neither fundamental risks nor superior information. If this is the case, investors may use EW portfolios to earn higher returns without bearing additional fundamental risk.

OA major concern when implementing EW strategies is the transaction cost. Compared with VW indexing which is eventually a buy-and-hold strategy, EW indexing relies on regular rebalancing thus will incur greater transaction costs. Whether the mispricing return premium delivered by the EW indexes could overcome the additional transaction costs becomes an empirical question. In this paper, we use effective spread and quoted spread as measures of transaction costs and confirm the robustness of EW strategies to transaction costs.

The rest of the essay is organized as follows. Sections 2 describes the model of the mispricing return premium and its impact on the performance of EW indexes, as well as simulation results that support our hypothesis. Section 3 presents empirical evidence based on a sample of member stocks of several major U.S. stock indexes. Section 4 concludes.

2. The Mispricing Return Premium and Its Impact on Portfolio Performance

2.1 The model of the mispricing return premium

As proposed by Brennan and Wang (2010), random mispricing could generate an upward return bias, denoted as the "mispricing return premium" due to Jensen's inequality. In a simplified version of the error-in-pricing framework, market price of a stock at the end of period t is denoted by P_t, while the unobservable efficient price denoted by P_t^*. $Z_t = P_t/P_t^*$ is used as a measure of the (relative) pricing error. It is assumed to have zero cross-sectional correlation and is independent from fundamental price, but could be serially correlated. The market price is assumed to be an unbiased estimate of efficient price so that $E[Z_t] = 1$. The natural logarithm of the pricing error, $z_t = \ln Z_t$, is assumed to be an independently and normally distributed random variable with possible serial correlation: $z_t \sim N(\mu, \sigma_{z,t}^2), \mu = -\frac{\sigma_{z,t}^2}{2}$, and $z_t = c + \rho z_{t-1} + \eta_t$. Its variance, $\sigma_{z,t}^2$, serves as the measure for the magnitude of the pricing error. For simplicity, the stock is assumed to pay no dividend. Based on above setting, the expected return of individual stock could be expressed as:

$$E\left[1+r_{i,t+1}\right]=E\left[\frac{P_{i,t+1}}{P_{i,t}}\right]=\left(1+r_{i,t}^{*}\right)E\left[\frac{Z_{i,t+1}}{Z_{i,t}}\right]=\left(1+r_{i,t}^{*}\right)e^{(1-\rho)\sigma_{Z_i}^{2}}, \tag{3.1}$$

where $r_{i,t}^{*}=P_{i,t}^{*}/P_{i,t-1}^{*}$ is the return of the fundamental price. As far as pricing errors exist and their correlation is less than one ($\rho<1$), the expected return will be higher than the fundamental return.

2.2 Returns of EW and VW portfolios

Under a simplified assumption of homogeneous fundamental return ($r_{i,t}^{*}=r_{t}^{*}$) and homogeneous magnitude of mispricing across stocks ($\sigma_{Z_i}^{2}=\sigma_{Z}^{2}$), expected return of an EW portfolio with *N* stocks is:

$$E[1+r_{EW}]=\sum_{i=1}^{N}\frac{(1+r_{t}^{*})e^{(1-\rho)\sigma_{Z}^{2}}}{N}=(1+r_{t}^{*})e^{(1-\rho)\sigma_{Z}^{2}}. \tag{3.2}$$

In the case of zero pricing error ($\sigma_{Z}^{2}=0$), expected return of the EW portfolio will be exactly the fundamental return of stocks ($E[1+r_{EW}]=1+r_{t}^{*}$), which is also the expected return of the VW portfolio in this case[41].

Given the presence of pricing error ($\sigma_{Z}^{2}>0$) and imperfect correlation in pricing errors ($\rho<1$), expected return of the EW portfolio will exceed the fundament return:

$$E[1+r_{EW}]=(1+r_{t}^{*})e^{(1-\rho)\sigma_{Z}^{2}}>1+r_{t}^{*}. \tag{3.3}$$

On the other hand, as mispricing of individual stocks largely cancel out at the portfolio level, the upward bias in VW portfolio returns should be much weaker, since it depends on the magnitude of portfolio-level mispricing. As suggested by Blume and Stambaugh (1983), returns on a VW strategy are virtually unbiased, since the buy-and-hold portfolios contain a "diversification" effect that removes virtually all bias.[42] Thus, mispricing return premium will generate a spread between the returns of an EW index and a VW index.

---

[41] Expected return of a VW portfolio: $E[1+r_{VW}]=\sum_{i=1}^{N}w_i(1+r_{t}^{*})=1+r_{t}^{*}$.

[42] This conclusion can be further illustrated by such a simple example. Suppose an investor hold a set of stocks that are fairly priced at the beginning and the end of a *T*-period investment horizon but experience

2.3 EW and VW portfolio returns based on simulated stock prices

Based on the framework presented above, we simulate 500 series of stock prices with random pricing errors across 348 periods for 100 iterations. Each stock is assumed to have only one share so its market value is equal to its stock price. Market values of all S&P 500 stocks at the end of 2011 are assigned to the simulated stocks to mimic the realistic cross-sectional variation in stock size. To simplify the simulation process and to eliminate complications from factors other than mispricing, we assume that all stocks zero average return, and that pricing errors have zero mean, zero autocorrelation, and homogeneous standard deviations across stocks and over time. Specifically, various pricing error volatilities are tested: σ_z = 0.5%, 1%, 2%, 3%, 5%, 10%, 15%, and 20%. We then compare returns of EW and VW portfolios based on the simulated prices. Panel A of Table 3.1 confirms that mispricing generates positive return premium for EW portfolios and the magnitude of this premium increases with the magnitude of pricing error measured by σ_z. More importantly, the mispricing return premium induced by Jensen's inequality has little effect on the returns of VW portfolios. Returns of the VW portfolio are always economically small and statistically insignificant. For instance, when an extremely large mispricing ($\sigma_z = 20\%$) is assumed in all 500 stocks, VW portfolio has a return of 0.03% in each period, which is statistically insignificant and much lower than 4.07%, the return of the EW portfolio in each period. Results are similar when 100 stocks are included into the portfolio (Panel B of Table 3.1). The results confirm that random mispricing could lead to a positive return spread between EW and VW indexes due to the upward return bias proposed by Brennan and Wang (2010).

Please insert Table 3.1 here

pricing errors during the holding period. Further assume that the magnitude of pricing errors is homogeneous and constant over time ($\sigma_{z_{i,t}}^2 = \sigma_z^2 > 0$), and that all stocks have fundamental return of zero in each period. EW strategies will benefit from the mispricing return premium, as regular portfolio rebalancing increase the weights of underpriced stocks and decrease the weight of overpriced stocks:

$$E[1 + HPR_{EW}] = e^{(1+(T-2)(1-\rho))\sigma_z^2}.$$

The VW strategy, however, will deliver a holding period return of zero. Therefore,

$$E[HPR_{EW} - HPR_{VW}] = e^{(1+(T-2)(1-\rho))\sigma_z^2} > 0.$$

3. Empirical Evidence

3.1 Sample of stocks

The empirical analysis of this paper is based on a total of 8,507 U.S. common stocks which were members of several major U.S. stock indexes over the sample period of January 1993 to December 2013. A total of eight indexes are used as testing platforms, including S&P 500 Composite Index, S&P 500 Growth Index, S&P 500 Value Index, S&P 400 Mid-Cap index, S&P 600 Small-Cap Index, Russell 1000 Index, Russell 2000 Index, and Russell 3000 Index. Constituents of the S&P index are obtained from Compustat and those of the Russell indices are obtained from Russell Investments. Monthly stock returns and characteristics are taken from CRSP. Intraday quote and trade data required in the estimation of inefficiency measures and transaction costs are taken from NYSE TAQ.[43] One-month T-bill rates and risk factors, including the three Fama-French (1993) factors and the momentum factor of Carhart (1997) are obtained from WRDS. Most variables used in this paper, including stock returns, are winsorized by 1% at each tail. Table 3.2 reports the summary statistics for these variables.

Please insert Table 3.2 here

3.2 Returns of EW strategies based on major U.S. equity indexes

We construct EW and VW portfolios based on members of each specific index. For example, to construct an EW version of the S&P 500 index, we assign equal weight to all member stocks of the S&P 500 index at the beginning of each rebalancing interval and estimate the equally-weighted average stock return over that month. Three rebalancing frequencies, monthly, quarterly and annually, are tested. For the VW version of the S&P 500 index, the weight of each member stock at the beginning of a month is proportional to the market capitalization of that stock at that time, and no rebalancing will happen until an

[43] Following Hasbrouck (1988) and Lee and Ready (1991), this paper assumes that trades are recorded five seconds late before 1998 and adjusts the time stamps accordingly. For the post-1998 period, the time stamps are assumed to be recorded correctly.

index addition or deletion. Time series of portfolio returns are obtained by concatenating the monthly EW or VW average returns of each portfolio and are then evaluated by a time-series regression on empirical risk factors. The time-series averages of the four-factor-adjusted (*FF3* and momentum adjusted) returns are estimated for both the EW portfolio and the VW portfolio, as well as the spread between the returns of both portfolios. Equally-weighted four-factor adjusted return ($\alpha_{FF4,EW}$) is also reported.

Please insert Table 3.3 here

Table 3.3 presents the four-factor alphas for the EW and VW versions of major U.S. stock indexes and their return spread, as well as annual portfolio turnovers and Sharpe ratios. Among the VW versions of the eight indexes, six of them have insignificant alphas and two of them have significantly negative alphas. Among the EW version of the eight indexes with monthly rebalancing, however, four of them have positive and significant alphas and none of them has negative alpha. As a result, seven out of the eight indexes shows positive and significant return spread between equal weighting and value weighting, and the EW-VW spread is marginally significant for the S&P 500 Growth index. For example, the EW S&P 500 index delivers a 0.17% monthly FF4-adjusted return compared with the VW S&P 500 index, while the EW Russell 2000 index provides a 0.39% monthly FF4-adjusted return compared to the VW Russell 2000 index. Quarterly rebalancing and annually rebalancing shows very similar results. Meanwhile, Sharpe ratio is higher for all the eight EW indexes. For example, VW S&P 500 index provides a Sharpe ratio of 0.13, while the EW index provides Sharpe ratios from 0.16 (monthly and quarterly rebalancing) to 0.17 (annual rebalancing). In sum, Table 3.3 confirms that EW indexes significantly outperform VW indexes.

3.3 Portfolio returns net of transaction costs

Since we are interested in whether investors should adopt EW strategies to earn the mispricing return premium, it is very important to examine whether this return spread is robust to transaction costs. Compare with VW indexing which is eventually a buy-and-hold strategy, EW indexing relies on regular rebalancing thus will incur greater transaction costs. For example, as presented in Table 3.3, the annual turnover ratios of the EW indexes are between 0.94 and 1.70 for monthly rebalancing, between 0.62 and 1.16 for quarterly rebalancing, and between 0.29 and 0.46 for annual rebalancing. Therefore, whether the mispricing return premium delivered by the EW indexes could overcome the additional transaction costs becomes an empirical question. We use either relative effective spread (*RES*) or relative quoted spread (*RQS*) during the contemporaneous month of the rebalancing to measure the round-trip transaction costs. The trade-weighted *RES* is estimated as two times the absolute distance between actual transaction price and the prevailing quote midpoint, scaled by the quote midpoint; the time-weighted *RQS* is estimated as the absolute distance between the bid price and the ask price, scaled by the quote midpoint. Both measures are widely used in the literature but they differ significantly in magnitudes. For example, Petersen and Fialkowski (1994) find that effective spreads are only half of the quoted spreads and that increase in the quoted spread results in only 10%-22% increase in the effective spreads. The quoted spread significantly overestimates the true transaction costs because many trades are executed inside the quoted spreads. Therefore, *RES* is used as the main proxy for transaction costs while *RQS* is only for robustness.

Please insert Table 3.4 here

Table 3.4 reports the four-factor adjusted alphas based on net returns where transaction costs are deducted for EW indexes[44]. Specifically, returns of the EW indexes are first estimated by quote-midpoint returns. Total transaction costs are then estimated based on the trading volume of each stock in each rebalancing and the corresponding *RES* or *RQS*. When transaction costs are measured by *RES*, net returns of EW indexes slightly decreases, but are still significantly higher than returns of VW indexes for all three rebalancing frequencies. For example, with quarterly rebalancing, EW S&P 500 index outperforms VW S&P 500 index by a monthly alpha of 0.17%, or over 2% annually; EW Russell 2000 index outperform VW Russell 2000 index by a monthly alpha of 0.37%, or over 5% annually. Thus the outperformance of the EW indexes are both statistically and

[44] Transaction costs for VW indexes are assumed to be zero, as the turnover ratio for VW indexes are almost zero. This would slightly underestimate the EW-VW spread.

economically significant even after deducting transaction costs. When transaction costs are measured by *RQS*, EW returns are noticeably lowered, but are still significantly higher than VW returns for most of the eight indexes. For example, with quarterly rebalancing, all EW indexes except for the S&P 500 Growth significantly outperform the VW indexes. With quarterly rebalancing, EW S&P 500 index outperforms VW S&P 500 index by a monthly alpha of 0.16%, or approximately 2% annually; EW Russell 2000 index outperform VW Russell 2000 index by a monthly alpha of 0.32%, or approximately 4.5% annually. Furthermore, Sharpe ratios are higher for most of the EW indexes, especially under annual rebalancing. Therefore, the outperformance of EW indexes is robust to the deduction of transaction costs[45].

3.4 Strategies based on price inefficiency

The above results are consistent with our hypothesis that the mispricing return premium could generate the return spread between EW and VW indexes. However, further evidence is needed to prove than this outperformance could indeed be explained by mispricing return premium. As the magnitude of mispricing return premium depends crucially on the magnitude of random pricing errors, an EW portfolio would outperform the corresponding VW portfolio only when the underlying stocks have pricing errors. Based on this concept, we repeat our tests based on subsample of stock sorted by the magnitude of mispricing.

As efficient price is, unfortunately, unobservable and difficult to estimate, it is impossible to precisely estimate σ_z^2 in the model proposed in Section 1. The use of empirical proxies of price inefficiency, which are usually noisy and controversial though, is a necessity. Following the literature, we assume that a fully-efficient stock price should follow a random walk, and we deviation of stock price from random as empirical proxy of mispricing. Specifically, we use Hasbrouck's (1993) pricing error variance as the main

[45] We only consider explicit transaction costs because the lacking of well-established methodology to estimate market impact cost. Therefore, our finding may not be robustness for a very large portfolio which could incur noticeable price impact during a rebalancing.

proxy for price inefficiency, and use daily first-order return autocorrelation for robustness tests.

The pricing error proposed by Hasbrouck (1993) measures the deviation between transaction prices and an implicit random walk process, which is assumed to be the efficient price. Specifically, the log transaction price, p_t, is defined as the efficient price, m_t, plus a transitory deviation, s_t:

$$p_t = m_t + s_t. \tag{4.4}$$

t indexes transaction time; m_t is defined as the expectation of the stock value given all available public information and is assumed to follow a random walk; s_t measures the deviation of transaction price from the efficient price. Specifically, s_t is assumed to be a zero-mean covariance-stationary stochastic process with variance of σ_s^2. Clearly, σ_s^2 inversely measures how closely the transaction price follows the efficient price, thus is used as a measure of price inefficiency.

Following Hasbrouck (1993), we estimate the *lower bound* for σ_s^2, denoted as V(s), using a five-lag vector autoregression (VAR) model based on intraday trade and quote data obtained from NYSE TAQ[46]. However, estimation of σ_s^2 is largely affected by the return volatility, making the comparison of σ_s^2 across stocks much less meaningful. To control for the cross-sectional difference in stock return variance, we follow Boehmer, Saar, and Yu (2005) and several other studies to normalize σ_s^2 by the variance of the log transaction prices, σ_p^2, to form a relative measure of inefficiency, σ_s^2/σ_p^2 (denoted as V(s)/V(p)). It represents the proportion of deviations from random walk in the total variability of the transaction return process and are used as the main proxy for inefficiency in this paper.

Please insert Table 3.5 here

At the beginning of each month, we sort all member stocks of a specific index by their V(s)/V(p) estimated in the previous month, then construct EW indexes and VW indexes based on the top 50% stocks and the bottom 50% stocks, respectively. Table 3.5 presents the four-factor-adjusted return spread between the EW and the VW indexes.

[46] See Appendix A for detailed estimation process.

Consistent with our hypothesis, the EW-VW spreads are positive and significant for all the eight indexes based on high pricing error stocks, while only two of the eight indexes based on stocks with low pricing errors show significant EW-VW spread. For example, the monthly four-factor adjusted EW-VW spread based on S&P 500 stocks with high pricing errors is 0.24% (t = 2.80), while the one based on stocks with low pricing errors is only 0.02% (t = 0.30). Other indexes show very similar patterns. The two exceptions, the S&P 600 Small-Cap index and the Russell 2000 index are both based on small stocks, which are more likely to have large pricing errors. That may explain why the EW-VW spreads based on low-pricing-error members of both indexes are still statistically significant. Moreover, our findings are not changed even when transaction costs are deducted from portfolio returns, as reported in Table 3.6.

Please insert Table 3.6 here

3.5 Alternative measure of mispricing

We use the absolute value of daily first-order return autocorrelation, $|AC(1)|$, as an alternative measure of mispricing in robustness tests. Compared with $V(\mathrm{s})/V(\mathrm{p})$, which needs intraday trade and quote data in the estimation process, daily autocorrelation is more intuitive and much easier to estimate, and is more widely used in the literature. Specifically, $|AC(1)|$ is estimated for each stock over each month by regressing daily returns on 1-day lagged returns. As reported by Table 3.7, autocorrelation generates qualitatively similar results. For example, with monthly rebalancing, the EW-VW spread is statistically significant for seven out of the eight indexes based on high-pricing-error stocks, while it is significant for only three out of the eight indexes based on low-pricing-error stocks. The magnitudes of the EW-VW spread are generally larger with higher autocorrelation. When transaction costs, measured by *RES*, are considered, the EW-VW spread with monthly rebalancing is statistically significant for five out of eight indexes based on high-pricing-error stocks, while it is significant for only one index based on low-pricing-error stocks. Therefore, our findings are robust to the use of autocorrelation as an alternative measure of mispricing.

Please insert Table 3.7 here

3.6 Other robustness tests

As EW indexes put relatively more weight on small stocks and small stocks tend to outperform in January, the outperformance of EW indexes could be a result of the January effect. This possibility is precluded by Panel A of Table 3.8, which shows that outperformance of the EW indexes is not affected when Januarys are excluded from the stock sample. When testing the index returns in subperiods, however, the results become slightly weaker. For example, as reported in Panel B of Table 3.8, the EW-VW return spreads (net of transaction costs; with quarterly rebalancing) have relatively small magnitudes and are not statistically significant for six out of the eight indexes over 1993-2000. The statistical significance also drops over 2001-2007 and 2008-2013, probably due to shortened sample period thus greater standard errors.

Please insert Table 3.8 here

3.7 Concerns about investment capacity

As EW strategies assign equal weights to stocks with large and small market capitalizations, smaller stocks may face a higher upward price pressure due to their limited investment capacities. In an extreme case when all the investors are using EW strategies, the market capitalizations of all stocks must be the same. This obviously implies overpricing in stocks with low fundamental value and underpricing in stocks with high fundamental value. VW strategy, in contrast, does not suffer from the limitation of investment capacity, as the weight of each stock is proportional to its market capitalization.

4. Conclusion

The puzzling outperformance of equally-weighted indexes has been documented for a long time, yet there has been no agreement about the underlying reasons. Despite such a noticeable outperformance, value-weighted indexing is still the mainstream in today's indexing sector. In this paper, we examine whether the upward return bias generated by Jensen's inequality could explain such a puzzling phenomenon, and find supporting evidence. For a wide range of U.S. stock indexes, EW indexes deliver better four-factor adjusted returns than VW indexes, and this outperformance concentrates in stocks with greater price inefficiency, as measured by deviation of stock prices from random walk. We also resolve the concern of higher transaction costs associated with the EW index, finding that the outperformance of EW indexes is statistically and economically significant even after deducting transaction costs. Findings in this paper not only provide new insight into the long-term debate on causes of the outperformance of the EW indexes, but also imply a potentially winning investment strategy. Of course, we acknowledge the limitation of investment capacity on the real-world implementation of the EW indexes.

References 3

Arnott, Robert. 2006. An Overwrought Orthodoxy. Institutional Investor (18 December): 36–41.

Blume, Marshall E. and Robert F. Stambaugh, 1983. Biases in computed returns. Journal of Financial Economics 12, 387-404.

Boehmer, Ekkehart, Gideon Saar, and Lei Yu, 2005. Lifting the Veil: An analysis of pre-trade transparency at the NYSE. Journal of Finance Vol. 60, Issue 2, 783-815.

Brennan, Michael J. and Ashley W. Wang, 2010. The mispricing return premium. Review of Financial Studies Vol. 23, No. 9, 3437-3468.

Grossman, Sanford J., and Joseph E. Stiglitz, 1980. On the impossibility of informationally efficient markets. American Economic Review, Vol. 70, No. 3, 393-408.

Hasbrouck, Joel, 1988. Trades, quotes, inventories and information. Journal of Financial Economics 22, 229-252.

Hasbrouck, Joel, 1993. Assessing the quality of a security market: A new approach to transaction-cost measurement. Review of Financial Studies Vol. 6, No. 1, 191-212.

Hsu, Jason C., 2006. Cap-weighted portfolios are sub-optimal portfolios. Journal of Investment Management Vol. 4, No. 3, 1-10.

Lee, Charles M. C., and Mark J. Ready, 1991. Inferring trade direction from intraday data. Journal of Finance Vol. 46, Issue 2, 733-746.

Newey, Whitney K., and Kenneth D. West, 1987. A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. Econometrica Vol. 55, No. 3, 703-708.

Perold, Andre F., 2007. Fundamentally flawed indexing. Financial Analysts Journal Vol. 63, No. 6, 31-37.

Siegel, Jeremy J. 2006. The ‘Noisy Market’ Hypothesis. Wall Street Journal (14 June): A14.

Table 3.1: Portfolio returns with different levels of simulated mispricing

Stochastic pricing errors are simulated for a number of stocks over 348 periods. Pricing error of each stock is assumed to follow a lognormal distribution with zero mean and certain volatility σ_z. Market values of the S&P 500 stocks at the end of year 2011 are assigned to the simulated stocks. To simplify the simulation process and to eliminate potential complications, all stocks are assumed to have zero return on their efficient price thus have no change in their intrinsic market values. The volatility of pricing errors is assumed to range from 0.5% to 20%. The time-series average of the equally-weighted and value-weighted portfolios returns based on simulated stock prices are reported. Panel A reports results for portfolios with 500 stocks. Panel B presents results for portfolios with only 100 stocks. The *t*-statistics are based on the time-series variation in portfolio returns over the 348 months and are adjusted by the Newey and West (1987) standard errors.

	Volatility of Pricing Error (σ_z)							
	0.5%	1%	2%	3%	5%	10%	15%	20%
Panel A: The portfolio has 500 stocks								
EW Port. Ret.	0.00%	0.01%	0.04%	0.09%	0.25%	1.00%	2.27%	4.07%
t-statistic	3.54	7.08	14.08	21.08	34.95	66.65	91.75	112.95
VW Port. Ret.	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.02%	0.03%
t-statistic	0.02	0.03	0.06	0.08	0.14	0.28	0.40	0.57
Panel B: The portfolio has 100 stocks								
EW Port. Ret.	0.00%	0.01%	0.04%	0.09%	0.25%	1.00%	2.27%	4.07%
t-statistic	1.57	3.16	6.33	9.43	15.70	30.01	42.54	52.55
VW Port. Ret.	0.00%	0.00%	0.00%	0.00%	0.01%	0.03%	0.08%	0.14%
t-statistic	0.02	0.06	0.12	0.18	0.29	0.58	0.86	1.16

Table 3.2: Summary Statistics

The monthly sample is based on 3,180 U.S. common stocks which were members of several major U.S. stock indexes over the period 1993-2013. r is the monthly close-to-close stock return minus risk-free rate. r^{MP} is the monthly quote-midpoint stock return minus risk-free rate. V(s) is the pricing error of Hasbrouck (1993) estimated over a month and V(s)/V(p) is the relative pricing error (scaled by standard deviation of log price, V(p), over that quarter). *MV* is the market value of a stock at the end of a month. *TO* is the monthly turnover ratio, measured by monthly trading volume scaled by total share outstanding. *RES* is monthly trade-weighted averages of relative effective spread. *RQS* is monthly time-weighted average of relative quoted spread.

	Mean	Std.Dev.	Min	P25	P50	P75	Max
Stock Daily Returns							
r	0.91%	13.94%	-43.09%	-5.98%	0.58%	7.26%	63.02%
r_{MP}	0.88%	13.91%	-42.79%	-5.95%	0.55%	7.21%	61.63%
Measures of Price Inefficiency							
V(s)/V(p) ($\times 10^{-2}$)	0.80%	2.62%	0.00%	0.01%	0.04%	0.33%	27.46%
$\|AC(1)\|$	0.132	0.104	0.002	0.051	0.108	0.188	0.564
Measure of Transaction Costs							
RES	0.69%	0.75%	0.06%	0.24%	0.47%	0.87%	10.39%
RQS	1.59%	1.52%	0.14%	0.69%	1.15%	1.93%	16.74%
Other Characteristics							
MV ($ billion)	4.11	16.83	0.00	0.28	0.68	2.13	626.55
TO (/quarter)	16.48%	16.84%	0.15%	5.42%	11.00%	21.15%	90.31%

Table 3.3: Return of EW and VW indexes

The monthly sample is based on 3,180 U.S. common stocks which were members of several major U.S. stock indexes over the period 1993-2013. The EW indexes are rebalanced at the beginning of each calendar month, quarter, or year, respectively. The FF-4 adjusted returns are reported for EW and VW indexes, as well as the return spread between both indexes. TO_{EW} is the annual turnover ratio of each EW index in each rebalancing frequency. S_{VW} and S_{EW} are the Sharpe ratios of VW and EW indexes, respectively. The *t*-statistics are based on the time-series variation in portfolio returns over the 252 months and are adjusted by the Newey and West (1987) standard errors.

Indexes			Monthly Rebalancing				Quarterly Rebalancing				Annual Rebalancing			
	α_{VW}	S_{VW}	α_{EW}	α_{EMV}	TO_{EW}	S_{EW}	α_{EW}	α_{EMV}	TO_{EW}	S_{EW}	α_{EW}	α_{EMV}	TO_{EW}	S_{EW}
SP500	-0.01%	0.13	0.16%	0.17%	0.94	0.17	0.17%	0.17%	0.62	0.17	0.12%	0.13%	0.32	0.17
t-statistic	-0.27		2.03	2.22			2.17	2.38			1.55	1.70		
SP500G	0.09%	0.13	0.25%	0.16%	1.29	0.16	0.26%	0.17%	0.93	0.16	0.20%	0.12%	0.31	0.16
t-statistic	1.57		2.58	1.87			2.65	1.95			2.27	1.46		
SP500V	-0.06%	0.13	0.15%	0.20%	1.12	0.16	0.15%	0.21%	0.80	0.17	0.11%	0.17%	0.29	0.16
t-statistic	-1.17		1.51	2.74			1.61	2.94			1.15	2.32		
SP400	0.06%	0.17	0.19%	0.14%	1.14	0.18	0.19%	0.14%	0.77	0.18	0.15%	0.10%	0.37	0.18
t-statistic	0.53		1.66	2.64			1.71	2.58			1.38	2.30		
SP600	-0.07%	0.15	0.15%	0.23%	1.34	0.17	0.15%	0.22%	0.89	0.17	0.12%	0.19%	0.40	0.17
t-statistic	-0.85		1.71	4.18			1.72	4.47			1.29	3.31		
R1000	-0.02%	0.13	0.20%	0.22%	1.15	0.17	0.22%	0.24%	0.78	0.17	0.20%	0.23%	0.35	0.18
t-statistic	-1.41		2.58	3.03			2.93	3.43			2.45	2.83		
R2000	-0.31%	0.10	0.08%	0.39%	1.70	0.14	0.11%	0.42%	1.16	0.15	0.11%	0.42%	0.46	0.16
t-statistic	-4.90		1.11	5.91			1.53	6.09			1.35	4.97		
R3000	-0.05%	0.13	0.12%	0.17%	1.43	0.15	0.15%	0.20%	0.95	0.16	0.15%	0.21%	0.42	0.17
t-statistic	-3.23		2.06	3.38			2.47	3.85			2.16	3.14		

Table 3.4: Return of EW and VW indexes after deducting transaction costs

The monthly sample is based on 3,180 U.S. common stocks which were members of several major U.S. stock indexes over the period 1993-2013. The EW indexes are rebalanced at the beginning of each calendar month, quarter, or year, respectively. The FF-4 adjusted returns are reported for EW and VW indexes, as well as the return spread between both indexes. S_{VW} and S_{EW} are the Sharpe ratios of VW and EW indexes, respectively. In Panel A, round-trip transaction costs are measured by trade-weighted relative effective spread, estimated as two times the absolute distance between actual transaction price and the prevailing quote midpoint, scaled by the quote midpoint. In Panel B, round-trip transaction costs are measured by time-weighted relative quoted spread, estimated as the absolute distance between the bid price and the ask price, scaled by the quote midpoint. The *t*-statistics are based on the time-series variation in portfolio returns over the 252 months and are adjusted by the Newey and West (1987) standard errors.

Indexes			Monthly Rebalancing			Quarterly Rebalancing			Annually Rebalancing		
	α_{VW}	S_{VW}	α_{EW}	α_{EMV}	S_{EW}	α_{EW}	α_{EMV}	S_{EW}	α_{EW}	α_{EMV}	S_{EW}
Panel A: TC is measured by effective spread.											
SP500	-0.01%	0.13	0.15%	0.15%	0.16	0.16%	0.17%	0.17	0.11%	0.12%	0.17
t-statistic	-0.27		1.89	2.06		2.07	2.27		1.51	1.65	
SP500G	0.09%	0.13	0.24%	0.15%	0.16	0.25%	0.16%	0.16	0.20%	0.11%	0.16
t-statistic	1.57		2.45	1.72		2.56	1.84		2.24	1.42	
SP500V	-0.06%	0.13	0.13%	0.19%	0.16	0.14%	0.20%	0.16	0.10%	0.16%	0.16
t-statistic	-1.17		1.38	2.57		1.51	2.80		1.11	2.27	
SP400	0.06%	0.17	0.17%	0.11%	0.17	0.18%	0.12%	0.18	0.15%	0.09%	0.18
t-statistic	0.53		1.48	2.26		1.59	2.33		1.32	2.16	
SP600	-0.07%	0.15	0.11%	0.19%	0.16	0.12%	0.19%	0.17	0.10%	0.17%	0.17
t-statistic	-0.85		1.27	3.57		1.41	4.05		1.15	3.18	
R1000	-0.02%	0.13	0.18%	0.20%	0.17	0.20%	0.23%	0.17	0.20%	0.22%	0.18
t-statistic	-1.41		2.35	2.78		2.76	3.26		2.39	2.78	
R2000	-0.31%	0.10	0.00%	0.31%	0.13	0.06%	0.37%	0.14	0.09%	0.40%	0.15
t-statistic	-4.90		0.05	4.98		0.81	5.55		1.12	4.86	
R3000	-0.05%	0.13	0.07%	0.12%	0.14	0.11%	0.16%	0.15	0.14%	0.19%	0.17
t-statistic	-3.23		1.17	2.42		1.90	3.24		1.98	2.97	
Panel B: TC is measured by quoted spread.											
SP500	-0.01%	0.13	0.13%	0.14%	0.16	0.15%	0.16%	0.16	0.11%	0.12%	0.17
t-statistic	-0.27		1.69	1.86		1.94	2.14		1.44	1.59	
SP500G	0.09%	0.13	0.22%	0.13%	0.15	0.23%	0.15%	0.15	0.19%	0.11%	0.16
t-statistic	1.57		2.25	1.50		2.41	1.67		2.20	1.37	
SP500V	-0.06%	0.13	0.11%	0.17%	0.16	0.13%	0.19%	0.16	0.10%	0.16%	0.16
t-statistic	-1.17		1.18	2.31		1.36	2.61		1.07	2.22	
SP400	0.06%	0.17	0.14%	0.09%	0.17	0.16%	0.10%	0.17	0.14%	0.08%	0.18
t-statistic	0.53		1.25	1.75		1.43	2.00		1.25	1.96	
SP600	-0.07%	0.15	0.07%	0.14%	0.16	0.09%	0.16%	0.16	0.09%	0.16%	0.17
t-statistic	-0.85		0.74	2.66		1.03	3.38		0.99	2.95	
R1000	-0.02%	0.13	0.16%	0.18%	0.16	0.19%	0.21%	0.17	0.19%	0.21%	0.18
t-statistic	-1.41		2.05	2.48		2.56	3.04		2.32	2.70	
R2000	-0.31%	0.10	-0.07%	0.24%	0.12	0.01%	0.32%	0.13	0.07%	0.38%	0.15
t-statistic	-4.90		-1.04	3.76		0.08	4.74		0.87	4.59	
R3000	-0.05%	0.13	0.01%	0.07%	0.14	0.07%	0.13%	0.15	0.12%	0.18%	0.16
t-statistic	-3.23		0.23	1.37		1.28	2.54		1.76	2.74	

Table 3.5: Return of EW and VW indexes with high or low mispricing

The monthly sample is based on 3,180 U.S. common stocks which were members of several major U.S. stock indexes over the period 1993-2013. The FF-4 adjusted return spreads between the EW and the VW indexes are reported. The EW indexes are rebalanced at the beginning of each calendar month, quarter, or year, respectively. Transaction costs are ignored. Mispricing is measured by Hasbrouck's pricing error variance scaled by return variance, $V(s)/V(p)$. At the beginning of each month, all member stocks of a specific index are sorted by their $V(s)/V(p)$ estimated in the previous month. EW indexes and VW indexes are then constructed based on the top 50% stocks and the bottom 50% stocks, respectively. The *t*-statistics are based on the time-series variation in portfolio returns over the 252 months and are adjusted by the Newey and West (1987) standard errors.

	Bottom 50% stocks by $V(s)/V(p)$			Top 50% stocks by $V(s)/V(p)$		
Rebalancing interval	Monthly	Quarterly	Annually	Monthly	Quarterly	Annually
S&P 500	0.05%	0.06%	0.02%	0.27%	0.27%	0.24%
t-statistic	0.79	0.95	0.30	3.18	3.29	2.80
S&P 500 Growth	0.04%	0.06%	0.00%	0.30%	0.29%	0.25%
t-statistic	0.45	0.58	0.02	2.58	2.54	2.31
S&P 500 Value	0.11%	0.12%	0.08%	0.28%	0.28%	0.24%
t-statistic	1.41	1.51	1.06	3.25	3.41	2.89
S&P 400	0.10%	0.10%	0.04%	0.15%	0.14%	0.11%
t-statistic	1.47	1.48	0.72	3.30	3.20	2.35
S&P 600	0.18%	0.17%	0.12%	0.12%	0.12%	0.15%
t-statistic	3.08	3.13	2.54	2.28	2.30	2.78
Russell 1000	0.08%	0.10%	0.08%	0.34%	0.36%	0.36%
t-statistic	1.05	1.42	1.04	4.85	5.18	5.14
Russell 2000	0.24%	0.26%	0.28%	0.31%	0.35%	0.37%
t-statistic	3.98	4.23	3.58	5.51	6.02	5.92
Russell 3000	-0.04%	-0.02%	-0.02%	0.22%	0.25%	0.26%
t-statistic	-0.69	-0.41	-0.36	2.69	3.04	2.94

Table 3.6: Net return of EW and VW indexes with high or low mispricing

The monthly sample is based on 3,180 U.S. common stocks which were members of several major U.S. stock indexes over the period 1993-2013. The FF-4 adjusted return spreads between the EW and the VW indexes are reported. The EW indexes are rebalanced at the beginning of each calendar month, quarter, or year, respectively. Mispricing is measured by Hasbrouck's pricing error variance scaled by return variance, $V(s)/V(p)$. In Panel A, round-trip transaction costs are measured by trade-weighted relative effective spread. In Panel B, round-trip transaction costs are measured by time-weighted relative quoted spread. The *t*-statistics are based on the time-series variation in portfolio returns over the 252 months and are adjusted by the Newey and West (1987) standard errors.

Rebalancing interval	Bottom 50% stocks by $V(s)/V(p)$			Top 50% stocks by $V(s)/V(p)$		
	Monthly	Quarterly	Annually	Monthly	Quarterly	Annually
Panel A: TC is measured by effective spread.						
S&P 500	0.03%	0.04%	0.01%	0.24%	0.24%	0.23%
t-statistic	0.37	0.57	0.19	2.79	2.95	2.70
S&P 500 Growth	0.01%	0.03%	-0.01%	0.26%	0.27%	0.24%
t-statistic	0.14	0.31	-0.06	2.31	2.31	2.25
S&P 500 Value	0.09%	0.09%	0.08%	0.25%	0.25%	0.23%
t-statistic	1.06	1.20	0.97	2.89	3.08	2.80
S&P 400	0.05%	0.06%	0.03%	0.10%	0.10%	0.10%
t-statistic	0.75	0.87	0.52	2.21	2.28	2.05
S&P 600	0.11%	0.11%	0.11%	0.03%	0.05%	0.13%
t-statistic	1.93	2.07	2.24	0.64	0.96	2.38
Russell 1000	0.04%	0.07%	0.07%	0.30%	0.32%	0.34%
t-statistic	0.55	0.96	0.92	4.27	4.69	5.01
Russell 2000	0.14%	0.18%	0.25%	0.16%	0.23%	0.33%
t-statistic	2.49	3.10	3.36	2.97	4.17	5.49
Russell 3000	-0.10%	-0.07%	-0.03%	0.11%	0.16%	0.23%
t-statistic	-1.74	-1.32	-0.61	1.38	2.07	2.67
Panel B: TC is measured by quoted spread.						
S&P 500	-0.03%	-0.01%	0.00%	0.19%	0.20%	0.21%
t-statistic	-0.36	-0.10	0.00	2.19	2.40	2.56
S&P 500 Growth	-0.04%	-0.02%	-0.02%	0.21%	0.22%	0.23%
t-statistic	-0.41	-0.19	-0.20	1.84	1.89	2.14
S&P 500 Value	0.03%	0.05%	0.06%	0.19%	0.21%	0.22%
t-statistic	0.38	0.58	0.81	2.27	2.51	2.66
S&P 400	-0.02%	-0.01%	0.01%	0.02%	0.04%	0.08%
t-statistic	-0.39	-0.10	0.19	0.49	0.82	1.63
S&P 600	0.01%	0.02%	0.08%	-0.08%	-0.04%	0.09%
t-statistic	0.15	0.42	1.68	-1.48	-0.80	1.81
Russell 1000	-0.02%	0.02%	0.05%	0.24%	0.27%	0.33%
t-statistic	-0.25	0.22	0.72	3.38	3.89	4.81
Russell 2000	0.03%	0.09%	0.23%	0.01%	0.10%	0.29%
t-statistic	0.56	1.50	2.94	0.10	1.79	4.72
Russell 3000	-0.17%	-0.13%	-0.05%	-0.01%	0.07%	0.20%
t-statistic	-3.02	-2.43	-0.92	-0.13	0.89	2.32

Table 3.7: Alternative measure of price inefficiency

The monthly sample is based on 3,180 U.S. common stocks which were members of several major U.S. stock indexes over the period 1993-2013. The FF-4 adjusted return spreads between the EW and the VW indexes are reported. The EW indexes are rebalanced at the beginning of each calendar month, quarter, or year, respectively. Mispricing is measured by the absolute value of daily first-order autocorrelation ($|AC(1)|$). At the beginning of each month, all member stocks of a specific index are sorted by their $|AC(1)|$ estimated in the previous month. EW indexes and VW indexes are then constructed based on the top 50% stocks and the bottom 50% stocks, respectively. In Panel A, returns are measured based on daily close prices and transaction costs are ignored. In Panel B, round-trip transaction costs are measured by trade-weighted relative effective spread. In Panel C, round-trip transaction costs are measured by time-weighted relative quoted spread. The *t*-statistics are based on the time-series variation in portfolio returns over the 252 months and are adjusted by the Newey and West (1987) standard errors.

Rebalancing interval	Bottom 50% stocks by $\|AC(1)\|$			Top 50% stocks by $\|AC(1)\|$		
	Monthly	Quarterly	Annually	Monthly	Quarterly	Annually
Panel A: TC is assumed to be zero.						
S&P 500	0.08%	0.08%	0.04%	0.27%	0.28%	0.24%
t-statistic	1.01	1.09	0.62	3.09	3.28	3.08
S&P 500 Growth	0.10%	0.11%	0.07%	0.23%	0.24%	0.22%
t-statistic	1.07	1.08	0.81	2.13	2.27	2.25
S&P 500 Value	0.16%	0.16%	0.12%	0.25%	0.26%	0.22%
t-statistic	1.74	1.83	1.35	2.91	3.10	2.82
S&P 400	0.18%	0.18%	0.14%	0.08%	0.08%	0.03%
t-statistic	3.06	3.14	2.44	1.46	1.30	0.64
S&P 600	0.24%	0.23%	0.20%	0.20%	0.20%	0.13%
t-statistic	3.89	4.04	3.45	3.67	3.84	2.46
Russell 1000	0.14%	0.17%	0.13%	0.31%	0.33%	0.30%
t-statistic	1.76	2.09	1.58	3.79	4.18	3.70
Russell 2000	0.40%	0.42%	0.38%	0.36%	0.39%	0.36%
t-statistic	5.30	5.36	4.50	6.04	6.42	5.40
Russell 3000	0.10%	0.12%	0.08%	0.26%	0.29%	0.26%
t-statistic	1.49	1.81	1.20	4.86	5.33	4.31

Table 3.7: Alternative measure of price inefficiency (continued)

Rebalancing interval	Bottom 50% stocks by $\|AC(1)\|$			Top 50% stocks by $\|AC(1)\|$		
	Monthly	Quarterly	Annually	Monthly	Quarterly	Annually
Panel B: TC is measured by effective spread.						
S&P 500	0.02%	0.03%	0.03%	0.21%	0.22%	0.23%
t-statistic	0.26	0.41	0.41	2.42	2.65	2.91
S&P 500 Growth	0.05%	0.06%	0.06%	0.18%	0.19%	0.21%
t-statistic	0.54	0.61	0.66	1.68	1.84	2.14
S&P 500 Value	0.10%	0.11%	0.11%	0.20%	0.21%	0.21%
t-statistic	1.16	1.30	1.21	2.27	2.47	2.65
S&P 400	0.09%	0.10%	0.11%	0.00%	0.00%	0.01%
t-statistic	1.60	1.85	2.09	-0.02	0.02	0.23
S&P 600	0.10%	0.11%	0.16%	0.06%	0.07%	0.09%
t-statistic	1.80	2.05	2.93	1.23	1.61	1.89
Russell 1000	0.07%	0.10%	0.11%	0.24%	0.26%	0.28%
t-statistic	0.89	1.30	1.37	2.95	3.41	3.52
Russell 2000	0.21%	0.26%	0.34%	0.17%	0.23%	0.32%
t-statistic	2.99	3.54	4.09	3.00	4.06	4.93
Russell 3000	-0.05%	0.00%	0.05%	0.11%	0.16%	0.23%
t-statistic	-0.72	-0.07	0.69	2.09	3.06	3.78
Panel C: TC is measured by quoted spread.						
S&P 500	-0.07%	-0.05%	0.01%	0.12%	0.15%	0.20%
t-statistic	-0.85	-0.63	0.10	1.42	1.71	2.67
S&P 500 Growth	-0.03%	-0.01%	0.04%	0.10%	0.12%	0.19%
t-statistic	-0.29	-0.12	0.44	0.95	1.14	1.97
S&P 500 Value	0.02%	0.03%	0.09%	0.11%	0.13%	0.19%
t-statistic	0.19	0.38	0.99	1.28	1.52	2.43
S&P 400	-0.04%	-0.01%	0.08%	-0.12%	-0.11%	-0.02%
t-statistic	-0.65	-0.16	1.54	-2.28	-1.94	-0.37
S&P 600	-0.08%	-0.05%	0.12%	-0.12%	-0.09%	0.05%
t-statistic	-1.38	-1.02	2.08	-2.47	-1.85	0.96
Russell 1000	-0.03%	0.01%	0.09%	0.14%	0.17%	0.26%
t-statistic	-0.35	0.14	1.07	1.67	2.18	3.23
Russell 2000	0.00%	0.08%	0.29%	-0.05%	0.04%	0.27%
t-statistic	0.03	1.01	3.39	-0.89	0.68	3.95
Russell 3000	-0.21%	-0.15%	0.01%	-0.07%	0.00%	0.18%
t-statistic	-3.38	-2.40	0.09	-1.34	0.06	2.96

Table 3.8: Other Robustness tests

The monthly sample is based on 3,180 U.S. common stocks which were members of several major U.S. stock indexes over the period 1993-2013. The EW indexes are rebalanced at the beginning of each calendar month, quarter, or year, respectively. Transaction costs are measured by effective spreads. The FF-4 adjusted returns are reported for EW and VW indexes (α_{EW} and α_{VW}), as well as the return spread between both indexes (α_{EMV}). S_{VW} and S_{EW} are the Sharpe ratios of VW and EW indexes, respectively. Panel A shows the results when Januarys are dropped. Panel B reports results for subperiods. The *t*-statistics are based on the time-series variation in portfolio returns over the 252 months and are adjusted by the Newey and West (1987) standard errors.

Indexes			Monthly Rebalancing			Quarterly Rebalancing			Annual Rebalancing		
	α_{VW}	S_{VW}	α_{EW}	α_{EMV}	S_{EW}	α_{EW}	α_{EMV}	S_{EW}	α_{EW}	α_{EMV}	S_{EW}
Panel A: Drop Januarys											
SP500	-0.01%	0.13	0.16%	0.17%	0.17	0.18%	0.19%	0.17	0.11%	0.12%	0.17
t-statistic	-0.39		2.03	2.30		2.20	2.51		1.47	1.68	
SP500G	0.09%	0.13	0.25%	0.16%	0.16	0.26%	0.17%	0.16	0.19%	0.10%	0.16
t-statistic	1.44		2.32	1.68		2.39	1.79		2.05	1.24	
SP500V	-0.04%	0.13	0.15%	0.19%	0.17	0.17%	0.21%	0.17	0.11%	0.15%	0.17
t-statistic	-0.76		1.67	2.83		1.84	3.11		1.21	2.24	
SP400	0.09%	0.17	0.22%	0.13%	0.18	0.23%	0.14%	0.18	0.18%	0.10%	0.19
t-statistic	0.85		1.98	2.36		2.12	2.48		1.73	2.19	
SP600	0.04%	0.15	0.17%	0.13%	0.17	0.18%	0.14%	0.18	0.15%	0.11%	0.18
t-statistic	0.44		1.76	2.51		1.98	2.99		1.54	2.17	
R1000	-0.03%	0.13	0.18%	0.21%	0.17	0.21%	0.24%	0.17	0.18%	0.21%	0.18
t-statistic	-1.46		2.42	2.85		2.85	3.32		2.34	2.70	
R2000	-0.23%	0.10	0.04%	0.27%	0.13	0.10%	0.33%	0.14	0.12%	0.36%	0.16
t-statistic	-3.63		0.54	4.56		1.46	5.48		1.71	5.22	
R3000	-0.05%	0.13	0.09%	0.14%	0.15	0.14%	0.19%	0.15	0.16%	0.21%	0.17
t-statistic	-2.68		1.58	2.71		2.46	3.71		2.48	3.49	

Table 3.8: Other Robustness tests (continued)

Indexes	1993-2000			2001-2007			2008-2013		
	α_{EMV}	S_{EW}	S_{VW}	α_{EMV}	S_{EW}	S_{VW}	α_{EMV}	S_{EW}	S_{VW}
Panel B: Subperiods									
SP500	0.09%	0.23	0.23	0.17%	0.12	0.02	0.11%	0.16	0.13
t-statistic	0.73			2.09			1.28		
SP500G	0.09%	0.24	0.24	0.23%	0.08	-0.01	0.06%	0.18	0.15
t-statistic	0.41			1.96			0.54		
SP500V	0.11%	0.20	0.23	0.13%	0.15	0.07	0.18%	0.15	0.10
t-statistic	0.75			1.78			2.61		
SP400	0.12%	0.19	0.22	0.15%	0.16	0.13	0.14%	0.18	0.15
t-statistic	1.33			1.74			1.49		
SP600	0.21%	0.15	0.15	0.14%	0.17	0.14	0.17%	0.18	0.15
t-statistic	2.34			1.88			2.75		
R1000	0.09%	0.21	0.22	0.29%	0.13	0.03	0.19%	0.18	0.13
t-statistic	0.71			3.78			1.77		
R2000	0.30%	0.11	0.09	0.32%	0.13	0.09	0.48%	0.18	0.12
t-statistic	3.34			2.99			4.40		
R3000	0.01%	0.14	0.21	0.19%	0.13	0.03	0.19%	0.18	0.13
t-statistic	0.09			2.55			2.41		