Fall 12-22-2010

THE EFFECT OF ETFS ON STOCK LIQUIDITY

Sophia J. W. Hamm
University of Pennsylvania, sophia.hamm@gmail.com

THE EFFECT OF ETFS ON STOCK LIQUIDITY

Abstract
This paper investigates the effect of the introduction of exchange-traded funds (ETFs) on the liquidity of individual stocks. Prior analytical studies suggest that uninformed investors strictly prefer trading ETFs to trading individual stocks in order to avoid trading against informed investors. As a result of uninformed investors' migration, the markets for individual stocks are predicted to become illiquid as ETFs become widely available. Using ETF trading and holdings data between 2002 and 2008, I test the hypothesis that the higher the percentage of a firm's shares held by ETFs, the higher the adverse selection cost to trade the firm's stock. I find that the availability of ETFs as an alternative trading option is positively associated with the adverse selection component of bid-ask spreads of stocks in ETFs. The positive association is shown to be stronger with ETFs holding more diversified portfolios of stocks, as uninformed investors' incentives to switch are stronger for more diversified ETFs. The increase in the adverse selection costs of individual stocks is transferred to the ETF-level adverse selection costs, and diversified ETFs are especially shown to suffer from illiquidity of their underlying stocks. This dynamics between stock-level and ETF-level adverse selection casts a doubt whether uninformed investors can avoid the adverse selection cost by trading ETFs as effectively as expected.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Accounting

First Advisor
Wayne R. Guay

Second Advisor
John E. Core

Third Advisor
Robert E. Verrecchia

Keywords
ETF, mutual fund, liquidity, information asymmetry, adverse selection

Subject Categories
Accounting

The Effect of ETFs on Stock Liquidity

Sophia Jihae Wee Hamm

A DISSERTATION

in

Accounting

For the Graduate Group in Managerial Science and Applied Economics

Presented to the Faculties of the University of Pennsylvania

In

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2010

Supervisor of Dissertation

____

Wayne Guay, Professor of Accounting

Graduate Group Chairperson

____

Eric Bradlow, Professor of Marketing, Statistics, and Education

Dissertation Committee

John Core, Professor of Accounting

Robert Verrecchia, Professor of Accounting

Richard Lambert, Professor of Accounting

DEDICATIONS

To my husband, Jihun, who has a great heart and a great mind,

and whose support has made the times of my doctoral work much easier.

To my parents and my brother, who have always loved and encouraged me.

And to my son, Jason, the joy of our lives.

ACKNOWLEDGEMENTS

I greatly appreciate the advice and guidance of my dissertation committee; John Core, Wayne Guay (chair), Richard Lambert, and Robert Verrecchia. I thank the faculty members, staff, and Ph.D. students at the Wharton Accounting Department and I am grateful to Christopher Armstrong, Brian Bushee, and Catherine Schrand for their valuable comments and discussions. This paper has benefited from the comments and suggestions made by workshop participants at the University of Pennsylvania, UCLA, Carnegie Mellon University, the City University of New York-Baruch, the University of Connecticut, London Business School, the Ohio State University, the University of Oregon, the University of Rochester, Temple University, the University of Texas at Dallas, Washington University at St. Louis, and the University of Utah. I gratefully acknowledge the financial support from the Wharton School, University of Pennsylvania.

ABSTRACT

THE EFFECT OF ETFS ON STOCK LIQUIDITY

Sophia Jihae Wee Hamm

Wayne R. Guay (Supervisor of Dissertation)

This paper investigates the effect of the introduction of exchange-traded funds (ETFs) on the liquidity of individual stocks. Prior analytical studies suggest that uninformed investors strictly prefer trading ETFs to trading individual stocks in order to avoid trading against informed investors. As a result of uninformed investors' migration, the markets for individual stocks are predicted to become illiquid as ETFs become widely available. Using ETF trading and holdings data between 2002 and 2008, I test the hypothesis that the higher the percentage of a firm's shares held by ETFs, the higher the adverse selection cost to trade the firm's stock. I find that the availability of ETFs as an alternative trading option is positively associated with the adverse selection component of bid-ask spreads of stocks in ETFs. The positive association is shown to be stronger with ETFs holding more diversified portfolios of stocks, as uninformed investors' incentives to switch are stronger for more diversified ETFs. The increase in the adverse selection costs of individual stocks is transferred to the ETF-level adverse selection costs, and diversified ETFs are especially shown to suffer from illiquidity of their underlying stocks. This dynamics between stock-level and ETF-level adverse selection casts a doubt whether uninformed investors can avoid the adverse selection cost by trading ETFs as effectively as expected.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. Introduction

This paper examines how the introduction of tradable baskets of stocks affects uninformed traders' incentives and adverse selection costs of trading individual stocks. The term "adverse selection" in this paper refers to the lambda (λ) in the Kyle [1985] model of the price formation process. Specifically, the lambda represents the illiquidity of a market for a security, and it determines how severely information asymmetry affects a security's price.

Kyle [1985] develops a model in which informed investors take advantage of uninformed investors and profit from trading on private information about the value of an asset. Subrahmanyam [1991] extends the model to a multi-asset economy setting where baskets of stocks are available for trading. Because private information about individual assets plays a smaller role at a portfolio level, the informational disadvantage assessed by a market maker is smaller in the market for baskets of stocks, leading to a lower adverse selection cost.[1] Therefore, uninformed traders should prefer trading baskets of stocks. Zingales [2009] goes so far as to argue that regulation prohibiting individual investors from investing in individual stocks and encouraging them to invest only in exchange-traded funds (ETFs) or mutual funds could function as the ultimate protection of uninformed investors. Even in the absence of this suggested regulation, however, uninformed investors who acknowledge the informational disadvantage of trading individual stocks are expected to switch to trading ETFs or mutual funds. The recent

[1] Empirical studies including Clarke and Shastri [2001], Hedge and McDermott [2004a], Berkman, Brailsford, and Frino [2005], and Frino, Kruk, and Lepone [2007] find that the adverse selection cost is smaller at a portfolio level.

popularity of ETFs and index funds seems to corroborate this benefit to uninformed investors.

A key implication of Kyle [1985] is that the existence of uninformed investors in the market for an individual stock is an important determinant of the adverse selection cost in the price formation process. Specifically, the fewer uninformed investors that participate in trading, the less liquid the market becomes. To test this prediction, I examine whether the adverse selection problem at the individual stock level becomes exacerbated as ETFs are introduced, and uninformed investors trade ETFs instead of individual stocks. An ETF provides an ideal setting to test the shift of incentives of uninformed investors because ETFs come with fewer of the shortcomings of mutual funds, such as minimum investment restrictions, high fees, agency costs, and inefficiency in trading cost distribution.[2] I use the percentage of a firm's shares held by ETFs as a proxy for alternative investment opportunities for uninformed investors, and hypothesize that the adverse selection cost of trading the stock increases with this measure.

Using 152,151 firm-quarter observations of 8,420 firms between 2002 and 2008, I find a positive association between the adverse selection component of bid-ask spreads of a stock and the percentage of a firm's shares held by ETFs, implying that the adverse selection cost increases as the market for an individual stock is deprived of noise trades due to ETFs. The association between the change in the adverse selection cost and the

[2] Among mutual funds, index mutual funds are the closest alternatives to ETFs with their low expense ratios and passive trading strategies. In supplemental analysis, I examine whether the percentage of shares held by index mutual funds has similar effects on adverse selection of individual stocks. I also examine shares held and traded by actively managed mutual funds, and find that these funds do not influence adverse selection costs at the individual stock level, suggesting that active mutual funds are viewed by the markets as informed institutional ownership.

change in the percentage of shares held by ETFs during the periods of introduction of new ETFs is also significant and positive. These results highlight the redistribution among investors of costs and benefits associated with discouraging uninformed investors from trading individual stocks in order to protect them against informed investors. I further show that ETFs composed of these individual stocks with increased adverse selection costs yield lower effectiveness in diversifying individual stocks' adverse selection costs, which partially negates the benefit of switching to ETFs.

Subrahmanyam [1991] identifies conditions under which uninformed traders prefer trading baskets of stocks. These conditions are closely related to the degree of diversification of these baskets of stocks. The more diversified a basket of stocks is, the more attractive it is to uninformed investors. Therefore, the market for an individual stock becomes more illiquid if a newly introduced basket holds a more diversified portfolio.[3] I find modest evidence that the positive association between the increase in the adverse selection cost and the increase in the percentage of shares held by ETFs is stronger when these ETFs duplicate diversified broad market indexes (as opposed to when they duplicate sector indexes).

I further test the implicit assumption used in the previous test: whether a portfolio with a higher degree of diversification is more likely to exhibit lower adverse selection costs. An ETF benefits from enhanced liquidity as liquidity is taken away from individual stocks the ETF holds, and this liquidity shift is predicted to be stronger for more diversified ETFs. Also, lambdas of individual stocks are more efficiently diversified

[3] In this paper, I often compare individual stocks and ETFs. I use the terms "individual stocks" or "underlying stocks" to refer to stocks held by ETFs, not individual stocks in general.

away if an ETF holds a diversified portfolio. Therefore, more diversified ETFs are expected to exhibit lower lambdas than less diversified ETFs. At the same time, however, as the liquidity shift is stronger for the more diversified ETFs, stocks held by these ETFs suffer higher illiquidity than stocks held by less diversified ETFs. Therefore, more diversified ETFs are expected to exhibit higher lambdas than less diversified ETFs. In summary, the sign of the association between diversification and portfolio-level adverse selection is an empirical issue given the inflow of uninformed investors to ETF trades.

I find that, *ceteris paribus*, ETF adverse selection costs decrease in the degree of diversification, as predicted by portfolio theory (Markowitz [1952]) and by the liquidity effect of the inflow of uninformed investors who prefer more diversified ETFs.[4] However, I also find that the association between an ETF lambda and lambdas of individual stocks is stronger for more diversified ETFs. This result contradicts portfolio theory, but confirms that the undiversified portion of adverse selection costs of stocks in more diversified ETFs is larger than expected, presumably due to the deprivation of noise trades in the market of individual stocks.

The remainder of this paper proceeds as follows: Section 2 briefly reviews the institutional background of ETFs, discusses relevant literature, and develops hypotheses. Section 3 explains variable measurements and describes the sample. Section 4 provides

[4] The classic portfolio theory about diversification of idiosyncratic risks (Markowitz [1952]) predicts that the portfolio-level risk relative to the collective asset-level risks is lower when a portfolio is more diversified. However, this prediction is less obvious when the formation of the portfolio itself causes uninformed investors to migrate to the portfolio (e.g., an ETF). I predict that this migration increases individual stocks' adverse selection and further increases the lambda of ETFs composed of these individual stocks.

empirical test designs and presents the main test results. Section 6 provides additional tests and results. Section 7 concludes the paper.

2. Background and Hypotheses Development

2.1 Institutional background on ETFs

In this section, I summarize institutional details related to ETFs, and explain why these investment portfolios are appropriate for testing hypotheses related to how trading behavior of uninformed investors influences adverse selection trading costs. ETFs have grown exponentially since the first ETF (SPDR S&P500:SPY) was introduced in 1993. By the end of 2008, 728 ETFs were traded on NYSE and AMEX. The sample used in this paper includes 133 sector ETFs and 140 diversified ETFs holding U.S. equities between 2002 and 2008. The total market value of these ETFs was approximately $99 billion in the first quarter of 2002. By the end of 2008, the total market value grew five times, and the market value of the oldest and the largest ETF (SPY) was approximately $83 billion. In 2008, the numbers of sector ETFs and diversified ETFs were similar, but the average of sector ETFs' market values was only 13% of the average of diversified ETFs' market values.

A comparison of ETFs with other types of funds highlights several distinct characteristics of ETFs. There are three legal forms of investment companies: mutual funds (open-end companies), closed-end companies, and unit investment trusts (UITs). ETFs are legally classified as open-end companies or UITs. Nevertheless, ETFs are neither exactly open-end funds nor UITs. Regardless of their legal classification, ETFs share certain characteristics of all three forms of investment companies. First, ETFs are traded in the secondary market, which is a feature of closed-end companies and UITs. Second, as with mutual funds, investors of ETFs buy and sell shares from management companies of funds. A fundamental difference between mutual funds and ETFs in this buying and selling process stems from the fact that ETF shares are created by management companies and sold in a block to authorized parties (market makers, specialist, and arbitrageurs). These authorized parties resell these ETF shares in the secondary markets to retail investors, whereas retail investors of mutual funds directly buy and sell from management companies. In addition, when purchasing and redeeming ETF shares from the ETF management companies, the authorized parties pay and receive in the form of a basket of stocks instead of cash. If there is any discrepancy between NAV and the market price of ETF shares at the market closing, the authorized parties can profit from it. If NAV is lower (higher) than the market price of an ETF, the authorized parties deposit (receive) the basket of stocks and receive (deposit) the ETF shares from the fund. This process is called the "creation (redemption) in-kind."[5] Therefore, ETFs can own between 0% and 100% of a given stock, unlike index futures and options that do not affect the supply of stock shares traded in the open market.[6]

ETFs have several advantages over mutual funds that are attractive to uniformed investors (where one can think about uniformed investors as individuals that are focused

---

[5] However, the arbitrage profit is not the main motivation for the authorized parties to create and redeem ETF shares, nor is the profit meaningfully large (Gastineau [2002]). Market makers adjust ETF inventories mainly due to the demand and supply of ETF shares in the secondary markets.

[6] This characteristic fits the pricing models where the supply of securities is not infinite. The supply of ETFs is limited and affects the supply of underlying stocks, whereas the supply of other derivatives such as index futures can be infinite. By the same token, I exclude short ETFs and ultra ETFs that use derivatives as underlying assets.

typically on investing their savings in low-cost, well-diversified portfolios). First, ETFs offer lower transaction costs. The annual expense ratios of ETFs are significantly lower than those of U.S. equity mutual funds, in part because they have no shareholder accounting.[7] In 2008, the average expense ratio was 0.55% (median 0.55%) for domestic equity ETFs, 1.05% (median 0.94%) for index mutual funds, and 1.41% (median 1.32%) for actively managed domestic equity mutual funds.[8][9] According to the 2009 investment company fact book by the investment company institute (ICI factbook [2009]), there is a clear trend of investors' preference to lower expense mutual funds. More than 100% of the new cash inflow has been to mutual funds with below-average expense ratios while mutual funds with above-average expense ratios have experienced cash outflows. Investors' distaste toward a high expense ratio partially explains the popularity of ETFs and index mutual funds. In addition, ETFs are neither front-loaded nor end-loaded, do not require a minimum investment amount, which is $3,000 (median) for major index mutual funds offered to individuals.[10]

Second, the prices of ETFs track their NAVs efficiently, and investors can freely buy or sell them intra-day at the market price. Even though closed-end funds are also

---

[7] Mutual fund investors pay fees (12b-1 fees) to cover ongoing expenses in addition to one-time sales loads, and more than 50% of these fees are used for shareholder services including shareholder accounting. Purchases and redemptions of mutual fund shares by investors incur purchases and sales of stocks in markets; these are considered to be transactions that require book-keeping, and fund accounting agents validate daily transactions (ICI factbook [2009]).

[8] These figures are from "Fund index recap" (May, 2009) by Lipper (www.lipperweb.com).

[9] In terms of expense ratio, index mutual funds and ETFs are comparable. As the large index mutual funds are the ones that offer notably lower expense ratios, the value-weighted average expense ratio of index mutual funds (0.23%) is smaller than that of ETFs (0.31%).

[10] The minimum initial investment requirement for the 29 largest U.S. equity index mutual funds ranges from $0 to $1 million with a median value of $3,000 and an average value of $92,000 (as of Oct. 2009, www.morningstar.com). The average minimum investment required for purchasing the 8 largest index mutual funds offered to institutions is $53 million. The majority of these large index mutual funds belong to the Vanguard fund family.

tradable intra-day, their market prices exhibit substantial deviations from NAV ("closed-end fund puzzle").[11] In contrast, the trading prices of ETFs approximate NAV most of the time, thanks in part to the expected daily arbitrage activity of the authorized parties after the close. Pontiff [1996, 2006] and Ackert and Tian [2000] conclude that ETFs exhibit lower deviations from NAVs; these papers attribute it to the lower arbitrage costs, as the holdings of ETFs are well disclosed to investors. In fact, all ETFs in the sample duplicate indexes whose market values are posted frequently during the trading hours. According to Ackert and Tian [2000], the discrepancy between the market price and NAV never exceeded 1% for the earlier ETFs (S&P500 SPDRs and MidCap SPDRs).

Third, trading costs of ETFs are more efficiently distributed among investors. Trading costs, including brokerage commissions, bid-ask spreads, and liquidity risk, are considered to be an additional fee ETF investors need to pay over and above the traditional mutual fund expenses. This notion stems from the fact that the investors themselves trade ETF shares in secondary markets, and presumably they do so more frequently than they do mutual funds. Gastineau [2002] notes that, contrary to this common belief that the trading costs are the disadvantage of ETFs, the trading costs (paid by those who actually trade) are the advantage of ETFs over mutual funds. It is well documented that long-term investors in mutual funds subsidize trading costs incurred by investors who frequently trade (Dickson, Shoven, and Sialm [2000], Christoffersen, Keim, and Musto [2007], and Guedj and Huang [2008]). This is because whenever mutual fund shares are created or redeemed at the market closing prices due to these frequent traders,

[11] Lee, Shleifer, and Thaler [1991] note that the norm is 10 to 20 percent discount from NAV and seek explanations from investor irrationality. Cherkes, Sagi, and Stanton [2009] attribute the discount to illiquidity of closed-end fund shares.

the adverse effect on the stock price due to mutual fund managers' activities of buying and selling underlying stocks only affects the remaining investors. The trading cost structure of ETFs forces frequent traders to bear their own trading costs, and therefore buy-and-hold investors are better off with ETFs than with mutual funds. The transaction costs that frequent traders of ETFs incur are not especially high, whereas trading mutual funds is usually subject to higher brokerage fees and wider spreads. This is because trading ETFs is considered to be the same as trading stocks, and the trading-cost feature innate in the bid-ask spreads is common to any assets traded intra-day. In addition, the trading cost of buying a basket of stocks as one security is considerably lower than the trading cost of buying the actual portfolio of stocks.[12]

In summary, ETFs offer uninformed investors several advantages over mutual funds with no obvious additional costs. ETFs' fees are significantly lower than mutual fund management fees, ETFs track their underlying stocks' net asset values more efficiently, and the trading costs are more efficiently distributed among investors based on the frequency of trades.

2.2 Adverse selection and market microstructure studies

This paper investigates whether the introduction of tradable baskets of stocks via ETFs has an impact on the adverse selection costs of individual assets as suggested by prior theories, using both firm-level and portfolio-level characteristics. In this section, I

[12] Bid-ask spreads of ETFs have the same characteristics as bid-ask spreads of stocks, and therefore I calculate the adverse selection component of ETF bid-ask spreads in the same way as I calculate the adverse selection component of stock bid-ask spreads.

review literature on the adverse selection costs of assets in general and theories to develop the hypothesis.

The market microstructure literature focuses on the price formation process through intra-day trading mechanisms. The main players in this price formation process are the (uninformed) market maker, informed traders, uninformed traders and noise traders. The market maker reacts to order flows submitted by these traders by adjusting bid-ask spreads. When adjusting bid-ask spreads, the market maker takes into consideration both his inventory and the expected informational advantage of informed traders.

The earlier stream of analytical studies emphasizes the inventory adjustment concern of the market maker (e.g. Amihud and Mendelson [1980] and Madhavan and Smidt [1993]). The focus of study was moved to information asymmetry in the price formation process by Glosten and Milgrom [1985] and Easley and O'Hara [1987]. Kyle [1985] presents a model of how private information is incorporated into trading processes when the market maker sets a price to clear orders, where the price anticipates the behavior of an informed trader. A number of studies extend the Kyle [1985] model with variations in the model specification.[13] The notion of adverse selection used in this paper is lambda (λ) as defined in the Kyle [1985] model. All the predictions in this paper are based on characteristics of lambda, such as the likelihood of informed trades and the uncertainty associated with an asset's value.

[13] Among others, Admati and Pfleiderer [1988], Holden and Subrahmanyam [1992], and Spiegel and Subrahmanyam [1992] extend Kyle [1985] by relaxing the condition of one informed investor or the condition of one uninformed trader. Verrecchia [2001] adds the precision of public disclosure as an additional parameter.

Later studies examine both the inventory concern and information asymmetry in the price formation process, and focus on the decomposition of a bid-ask spread into multiple components.[14] However, most of these analyses are limited to an analysis of an individual security. A few exceptions that incorporate the notion of diversification into the price formation process model include Gammill and Perold [1989], which attributes the success of index futures to the advantage of higher liquidity at a portfolio level, and Subrahmanyam [1991], which extends the Kyle [1985] model to the trading process of a tradable basket of stocks and presents conditions in which informed traders and uninformed traders choose between trading individual stocks and trading a basket of stocks. The analytical study by Gorton and Pennacchi [1993] provides a setting where the introduction of a security based on minimum-variance composites eliminates all liquidity trades at an individual stock level. The main hypothesis of this paper is based on these studies. I hypothesize that the introduction of tradable baskets of stocks has an impact on the adverse selection costs of individual assets, and test the hypothesis using ETF data.

Empirical studies that test predictions directly or indirectly derived from Subrahmanyam [1991] are not unprecedented. Most of these studies use index futures as the appropriate proxy for baskets of stocks, and investigate the difference in liquidity and adverse selection between underlying stocks of indexes and the index futures. Some attribute the cash-future basis to the difference in liquidity and adverse selection. Berkman, Brailsford, and Frino [2005] finds that adverse selection costs are lower for

---

[14] Lee and Ready [1991] suggests a method to determine whether a trading is driven by buy or sell order flows. Hasbrouck [1991], Foster and Viswanathan [1993], Lin, Sanger, and Booth [1995], Huang and Stoll [1997], Madhavan, Richardson, and Roomans [1997], Glosten and Harris [1998], and Stoll [2000] present modifications of analytical methods to decompose bid-ask spreads.

index futures than they are for underlying stocks, using the FTSE100 stock index futures contract and its underlying stocks traded on the London Stock Exchange. Frino, Kruk, and Lepone [2007] finds that, in Australian markets, futures trades are uninformed, implying low adverse selection in futures trading. Roll, Schwartz and Subrahmanyam [2007] empirically finds a positive association between the cash-future basis of NYSE composite index futures and liquidity costs of stocks on NYSE proxied by bid-ask spreads. This study focuses on the lead and lag relation between the basis and bid-ask spreads. As the most direct test of Subrahmanyam [1991] among studies using index futures, Jegadeesh and Subrahmanyam [1993] shows an increase in bid-ask spreads of S&P 500 stocks after the introduction of the S&P 500 futures. However, the evidence is largely descriptive and based on a small sample.

Some studies use tradable funds, such as ETFs or closed-end funds, to analyze portfolio-level adverse selection. Hasbrouck [2003] examines the intraday data of ETFs and futures of three indexes (S&P 500, S&P 400, and Nasdaq 100), but the focus is on whether trades of these ETFs convey information that leads the price of index futures and vice versa. Clarke and Shastri [2001] finds that adverse selection for closed-end funds is lower than the weighted average of the adverse selection costs of underlying stocks. Hedge and McDermott [2004a] predicts that individual stocks become illiquid as ETFs become available. However, as a testable hypothesis for this prediction, the study tests whether the liquidity costs of ETFs are in general lower than liquidity costs of individual stocks, instead of directly testing the effect of ETFs on underlying stocks' liquidity. The study finds that the liquidity costs of two ETFs, DIA and Q's, are lower than their component stocks' liquidity costs and attribute the difference to lower adverse selection at an ETF level. Neal and Wheatley [1998] investigates the adverse selection components of the bid-ask spreads of 17 closed-end mutual funds. The study implicitly assumes that information asymmetry at the fund level should not exist as long as the net asset value of the underlying portfolio is frequently disclosed. Therefore, the observed adverse selection cost is attributed to mispricing, fund expense, and volume. These predictions, however, are not empirically confirmed. These prior studies that investigate ETFs' adverse selection costs do not directly address nor test the effect of ETFs on individual assets' adverse selection costs. This paper provides cross-sectional tests using a comprehensive sample of U.S. equity ETFs supplemented by the sample of U.S. equity mutual funds. The main purpose of this paper is to directly examine the effect of tradable baskets of stocks on the adverse selection costs of individual stocks.

In summary, this paper contributes to the existing studies in two aspects: it uses ETF data as the most suitable setting for testing adverse selection hypotheses derived from Subrahmanyam [1991], and it provides a direct test of the effect of a tradable baskets of stocks on the underlying stocks' adverse selection costs. In the next section, I provide hypotheses based on Subrahmanyam [1991]'s predictions, combined with the classical portfolio theory by Markowitz [1952] and the price formation process model by Kyle [1985].

2.3 Hypotheses development

I adopt two of Subrahmanyam [1991]'s predictions in this paper: (1) The benefit to uninformed investors from trading a basket of stocks instead of individual stocks increases as the basket becomes more diversified; and (2) The introduction of a tradable basket of stocks affects the demographics of investors, and therefore affects the liquidity of markets for these stocks. Based on these predictions, I develop a hypothesis that the introduction of a tradable basket of stocks makes the underlying securities less liquid, and that the more diversified the basket, the stronger this liquidity effect.

Subrahmanyam [1991] analytically demonstrates that when informed and uninformed investors have discretion over choosing between trading a pre-weighted basket of stocks as a single security versus individually trading the exact same composition of stocks, there exists a Nash equilibrium in which uninformed investors prefer to trade the basket security. As discussed below, there are fewer informed investors at a portfolio level. Therefore, uninformed investors prefer trading baskets because the adverse selection cost is smaller at a portfolio level than at an individual stock level.[15]

Informed investors' preferences are also important, as their preferences affect uninformed investors' decisions as to the kinds of tradable baskets to which they migrate. As it relates to these baskets, however, informed investors' preferences are less obvious than uninformed investors'. Investors who are privately informed about the idiosyncratic component of individual stocks ("security-specific informed investors") prefer to trade

[15] Easley, Engle, O'Hara, and Wu [2008] finds that the arrival rate of uninformed investors is negatively associated with the arrival rate of informed traders in the past.

individual stocks versus a pre-weighted basket of stocks. This preference is stronger when the basket of stocks is well-diversified and has a larger number of stocks.

Even though informed investors are, as described above, reluctant to trade baskets of stocks due to their loss of informational advantage, trading baskets proffers the benefit of both greater diversification of idiosyncratic risks and more liquidity. Therefore, some informed traders may reshuffle their resources and opt to acquire private information about the systematic component of a basket of stocks, instead of private information about security-specific components of the underlying stocks. An informed trader who gathers private information about the systematic component ("factor-informed investor") is better off trading a basket of stocks than trading individual stocks. Subrahmanyam [1991] predicts that a well-diversified basket of stocks attracts fewer security-specific informed traders, and more factor-informed traders. This prediction, however, needs to be interpreted cautiously. Consider two tradable baskets of stocks: a broader market index fund and a sector fund. The price of the former is largely determined by market risk because its idiosyncratic components are diversified away. The sector fund's price is also determined by market risk, but a sector fund diversifies away less idiosyncratic risk. For an investor who is informed about the market-wide systematic component, it is more profitable to trade a broader market index fund than a sector fund. Moreover, because security-specific informed investors are less likely to trade a broader-market index fund than a sector fund, there is less competition among informed traders. Therefore, conditional on an investor's possession of private information about a market-wide component, a broader market index fund is more attractive to the investor than a sector fund.

If we consider the costs and benefits for investors to acquire private information about a systematic factor, however, a sector fund can be a more attractive option. An informed investor who acquires private information about a certain industry can also be described as a factor-informed investor in the sense that the macro variable is a systematic factor of the industry. I predict that it is less costly for a security-specific informed investor to be a factor-informed investor of the industry to which the security belongs than to be a factor-informed investor of the market portfolio. Therefore, I predict that the likelihood of informed trade is higher for sector funds than for diversified, broader market index funds.

In summary, when a basket of stocks becomes available for trading, uninformed investors prefer trading the basket of stocks to trading an individual stock. I predict this migration of uninformed investors causes the adverse selection components of stock bid-ask spreads to increase. Security-specific informed traders choose to either continue to trade individual stocks or reallocate resources to acquire private information about the systematic component of the basket of stocks and trade the basket. The more diversified the newly available basket of stocks, the more likely the security-specific investors will continue to trade individual stocks. When they decide to become factor-specific informed traders, it is easier for informed traders to switch to trading sector funds than to trading well-diversified funds. As a result, uninformed traders assess the likelihood of informed trades to be smaller in trading diversified funds than in trading sector funds, and therefore are more likely to seek out diversified funds. This leads to a greater increase in adverse selection for stocks held by diversified portfolios than for stocks held by sector portfolios. These predictions lead to the following hypotheses:

Hypothesis 1a: The introduction of a basket of stocks makes the adverse selection problem of underlying stocks more severe. As a result, a stock with a higher percentage of shares traded as part of a basket of stocks exhibits greater adverse selection.

Hypothesis 1b: The increase in adverse selection is more pronounced when a basket of stocks is a diversified portfolio rather than a sector portfolio.

Uninformed investors' inclination toward more diversified portfolios can be explained in the context of the traditional portfolio theory (Markowitz [1952]). First, note that Kyle [1995]'s notion of adverse selection (λ) is a function of the number of informed investors (N), the precision of cash flow conditional on public information ($h+n$), and the precision of noise trading (t) as the following (Appendix A.1):

$$\lambda = \frac{\sqrt{Nt}}{N+1}\sqrt{\frac{1}{h+n}} \tag{1}$$

There are distinct trading markets, investor groups, and market makers for ETFs, and an ETF's closing price is not automatically the net asset value of the portfolio it holds. Therefore, an ETF's adverse selection is not exactly the weighted average of λs of the stocks held by the ETF. The theoretical value of an ETF's adverse selection is close to the portfolio-level adverse selection cost calculated as if the portfolio is a single security.

Appendix A.2 shows that portfolio-level adverse selection (λ_{pf}) can be approximated as follows: [16]

$$\lambda_{pf} = \left(\frac{\sqrt{N_{pf} t_{pf}}}{N_{pf} + 1} \right) \sqrt{\frac{1}{e^T V^{-1} e}}, \tag{2}$$

where N_{pf} is the number of informed traders trading the portfolio, t_{pf} is the precision of noise trading, e is a J x 1 vector of scalar ones and V is the J x J covariance matrix of values of J stocks in the portfolio, conditional on the public disclosure quality.

It is analytically straightforward to show that λ_{pf} is determined by three factors: the likelihood of informed trades determined by N_{pf} and t_{pf}, the diversification benefit determined by the off-diagonal elements of V, and the base-line uncertainty of underlying stocks determined by the diagonal elements of V.[17] Given that uninformed investors of individual stocks migrate to more diversified funds and informed investors prefer sector funds to diversified funds, the likelihood of informed trade is expected to be lower for more diversified portfolios. Therefore, λ_{pf} is predicted to be lower for more diversified portfolios. The diversification benefit is also higher for more diversified portfolios. That is, *ceteris paribus*, a sector ETF's lambda is more closely associated with lambdas of underlying stocks, as there is less diversification. On the other hand, a well-diversified ETF's lambda is weakly associated with lambdas of individual stocks, and therefore is smaller than explained by these stock lambdas. In sum, the first two determinants of λ_{pf}

[16] Eq. (2) is an approximation because it assumes (for tractability) optimal weighting in the portfolio. Obviously, ETFs are not optimally weighted – they mostly mimic popular indices such as the S&P 500 or Russell indices.

[17] See Appendices A.3-A.6 for determinants of λ_{pf} .

lead to the prediction that λ_{pf} is lower for more diversified portfolios than for less diversified ones.

The third determinant of λ_{pf}, the base-line uncertainty of a portfolio, however, leads to the opposite prediction when combined with the liquidity effect hypothesized in the paper. In principle, λ_{pf} is not directly determined by underlying stocks' adverse selection costs (λ), but by the diagonal elements of V (underlying stocks' uncertainties) that eventually determine underlying stocks' adverse selection costs (λ). There is no evidence that diversified funds hold stocks that manifest higher or smaller uncertainty than sector funds. In other words, the cross-sectional variation among ETFs in the diagonal elements of V (and stock lambdas) are not by the cross-sectional variation in the degree of diversification of ETFs (except that, as described above, they are more likely to be cancelled out by the off-diagonal elements of V if an ETF is more diversified). According to the first two hypotheses in the paper, however, lambdas of individual stocks increase as stocks are incorporated into ETFs. The increase is greater as the ETFs become more diversified. The increased lambdas of individual stocks are larger than explained by the diagonal elements of V, and therefore a larger portion of the ETF lambda gets left undiversified. This leads to the prediction that λ_{pf} is higher for more diversified portfolios.

Based on these predictions, I provide two conflicting hypotheses as the following:

Hypothesis 2a: In equilibrium, a portfolio that offers more diversification by investing in a broader market index attracts fewer informed traders and more uninformed traders, and exhibits less adverse selection.

Hypothesis 2b: In equilibrium, a portfolio that offers more diversification by investing in a broader market index holds stocks with higher adverse selection costs, and exhibits more adverse selection.

It is ultimately an empirical issue to determine the sign of the association between diversification and the portfolio-level adverse selection. However, the above conflicting hypotheses are based on distinct explanations. Therefore, I attempt to test whether the effect of both informed trades and the diversification benefit on ETF lambdas is different from the effect of increased lambdas of underlying stocks. In the next section, I provide variable definitions, the description of data, and the empirical test design.

3. Variable Measurement and Data Description

In this section, I describe how I measure variables and the dataset used to construct the sample. Detailed descriptions are provided on the measurement of the adverse selection component of bid-ask spreads, the percentage of a firm's shares held by ETFs (which is the main test variable), the timing of variable measurement, and control variables.

3.1. Variable measurement

Adverse Selection:

I estimate the adverse selection component of the bid-ask spread (*LAMBDA*) using the method suggested by Madhavan, Richardson, and Roomans [1997]. The Madhavan et al. [1997] measure of *LAMBDA* is slightly modified to be used in cross-sectional analyses (Armstrong, Core, Taylor, and Verrecchia [2009]) with adjustment of the effect of the magnitude of a price. This measure of *LAMBDA* is estimated in two-stage regressions using intra-day trade and quote data gathered from the Trades and Automated Quotes (TAQ) database. I gather TAQ data for the periods between 2002 and 2008 for all actively traded firms with required Compustat and CRSP data available. First, trades and quotes are matched following the Lee and Ready [1991] method that determines whether a trade is buyer-initiated or seller-initiated. The sign of a trade, x_t, is set to be +1 if the trade is buyer-initiated and -1 if seller-initiated. Using these signs of trades, I estimate the firm-specific auto-regressive coefficient ρ from the following regression in each month:

$$x_t = \rho x_{t-1} + e_t \tag{3}$$

Then, in the following second-stage regression, adverse selection (λ) is estimated as the sensitivity of the change in the mid-point of bid and ask prices, deflated by the previous trade's mid-point of bid and ask prices, to the residual of the first-stage regression.

$$(p_{t-} p_{t-1})/p_{t-1} = \varphi(x_t - x_{t-1}) + \lambda (x_t - \rho x_{t-1}) + u_t, \tag{4}$$

where φ is interpreted as the dealer cost component of bid-ask spreads (the dealer's compensation for risk-bearing, transaction, and inventory holdings) and λ as the adverse selection cost component of bid-ask spreads. This λ is used as the main variable *LAMBDA* in this paper. *LAMBDA* is bounded by zero and one, but the actual data often produce a *LAMBDA* that is negative or larger than one. I exclude these out-of-range observations from the analysis based on the assumption that these observations are statistical outliers and do not threaten the validity of the Madhavan et al. [1997] method of estimating the adverse selection component of bid-ask spreads.[18] Finally, *LAMBDA* is multiplied by 100 and expressed as a percentage.

The percentage of shares outstanding held by ETFs:

The Thomson Financial (S12) database provides information on stocks held by ETFs and mutual funds. I gather information on stocks held by all U.S. equity ETFs between 2002 and 2008 and calculate the daily percentage of each firm's shares held by

[18] The study by Henker and Wang [2006] argues that matching trades and quotes assuming a one second delay between them is more suitable than Lee and Ready [1991]'s "five-second rule" to match quotes and trades. They find that 53% of the adverse selection component estimated using the Huang and Stoll [1997] method for S&P 500 firms in 1999 are negative and significant, and the percentage is decreased to 14% if the "one-second rule" is used. For the sample used in this paper, I use the Madhavan et al. [1997] method combined with Lee and Ready [1991]'s "five-second rule" and find only 1% of the firm-month lambdas are negative. Henker and Wang [2006] also notes that the Huang and Stoll [1997] method is less widely used than the Madhavan et al. [1997] method due to the high percentage of negative lambdas, implying that it is less of an issue with Madhavan et al. [1997]. Nevertheless, I match trades and quotes with one second suggested by Henker and Wang [2006] to see if the "five-second rule" compared to the "one-second rule" poses any concern. I find that 1.2% of the observations estimated by the "one-second rule" have negative lambdas. Although approximately 20% of negative lambdas from the "five-second rule" can be replaced by positive lambdas from the "one-second rule", I do not combine lambdas from these two tick rules to maintain consistency of the sample. Instead, I exclude those negative lambda observations from the sample. The results do not change if I replace negative lambda observations from the "five-second rule" with positive lambda observations from the "one-second rule", or replace them with zeros. In the ETF sample, the larger portion of ETF-month observations (23.5%) yields negative lambdas than in the firm sample. I examine negative lambdas for ETFs in 2005 and similar sized random sample of positive lambdas in 2005 to see the significance of lambdas, and find that these negative lambdas are all insignificant whereas 91% of positive lambdas are significant. I exclude negative lambda observations from the ETF sample.

ETFs (*ETF_%HLD*). Firms that are not held by any ETF in the sample periods are also included in the sample with *ETF_%HLD*=0. The final sample includes 152,151 firm-quarter *ETF_%HLD* observations of 8,420 firms. Because the S12 database provides holdings data that are reported quarterly in most cases, I assume that holdings on the reporting date are applied to dates between the reporting date and the latest date among the following: the prior reporting date, 365 days prior to the reporting date, or the ETF inception date. *ETF_%HLD* is the quarterly average of daily percentages of a firm's shares held by ETFs.

The percentage of shares outstanding held by mutual funds:

Index mutual funds provide uninformed investors with similar benefits of passive index trading with relatively low costs. To examine whether index mutual funds also decrease liquidity in the markets of their underlying stocks, I construct *MFID_%HLD*, the percentage of a firm's shares held by index mutual funds. *MFID_%HLD* is calculated in the same manner as *ETF_%HLD*, using funds in the S12 database that are manually identified as index mutual funds from their fund names. To use as a control for the role of actively managed mutual funds, *MFAM_%HLD* is calculated likewise as the percentage of a firm's shares held by actively managed mutual funds.

Post-period and pre-period for change variables:

In addition to firm-quarter observations, I use change variables between a pre-period and a post-period in order to examine the effect of changes in explanatory variables on the change in firms' adverse selection costs. I define an event period as a quarter in which new ETFs are introduced, set the timing of this quarter to be $t=0$, and define the pre-period as quarters from $t=-k$ to $t=-1$, where k varies from 1 to 4. Likewise, the post-period is defined as quarters from $t=1$ to $t=k$, where $k=1,..,4$. I calculate change variables as the post-period averages minus the pre-period averages.

Control variables:

To determine the control variables for the regression of *LAMBDA*, I again turn to the theoretical definition of adverse selection by Kyle [1985] described in eq. (1) of the previous section. In this model, the adverse selection cost is defined as $\lambda = \frac{\sqrt{Nt}}{N+1}\sqrt{\frac{1}{h+n}}$ where N is the number of informed investors, t is the precision (reciprocal of variance) of noise trades, h is the precision of a stock's expected value, and n is the precision of public information. Therefore, control variables are categorized, though not mutually exclusively, into variables related to the number of informed traders and noise traders, those related to the uncertainties about the firm value, and those related to the informational environment.[19]

First, *SIZE* and *BTM* determine adverse selection by affecting relatively long-term, assessed uncertainties about firm values. They also affect firms' information environments. *SIZE* is the log of one plus market value (CRSP price * shares), calculated

[19] Brennan and Subrahmanyam [1998] use three aspects of the lambda in the Kyle [1985] model to control for firm characteristics. They control for the variance of the asset payoff by the return volatility, the precision of private information by the number of analysts, and the size of noise trades by the size of the firm. Hedge and McDermott [2004b] also use the Kyle [1985] lambda to find out firm characteristics that determine adverse selection.

daily and averaged over the quarter. *BTM* is the book-to-market assets ratio calculated as Total assets / (Total assets – Book value of equity + Market value of equity) on the quarter-end date. Total assets and the book value of equity are from the Compustat quarterly database and the market value of equity is calculated using the CRSP database. Trading volume is generally higher for large firms, and stocks with high trading volume are considered to be more liquid. *VOLUME* is turnover calculated as the log of the ratio of one plus the quarterly average of daily CRSP trading volume and one plus the number of shares outstanding.

The return volatility and absolute abnormal returns indicate higher uncertainties about firm values and inferior public disclosure. *STDRET* is the annualized time-series standard deviation of CRSP daily returns, calculated each quarter. *ABS_ABNRET* is the absolute abnormal returns in each quarter, calculated as the absolute value of the difference between the cumulative return over the quarter and the cumulative value-weighted market return over the quarter. *PRICE* is the log of one plus the quarterly average of daily CRSP prices. The number of analysts proxies for the superior informational environment and a larger firm size. *ANALYST* is the log of one plus the quarterly average number of analysts from I/B/E/S. The percentage of institutional ownership and the number of institutions proxy for the likelihood of informed trades, firm sizes, and the competition among informed investors. *INST_%* is the percentage of institutional ownership and *INST_N* is the log of one plus the quarterly number of institutional owners, calculated with the Thomson Financial Spectrum (S34) data.

Additional Control Variables:

With the wide availability of computers, the frequency of intra-day trades, especially small trades, has notably increased in the past several years. The average daily number of trades has increased from 513 ($12 million) in the first quarter of 2002 to 6,702 ($35 million) in the last quarter of 2008. A trade is categorized to be small if it is smaller than $5,000. The percentage of small trades was 61% (39% in dollar volume) in the first quarter of 2001 and 89% (70% in dollar volume) in the last quarter of 2008. As stock lambdas can be affected by the macro-trend of the volume of small trades in the market, I include the quarterly average of daily percentages of intraday small trades (*PST*) as an additional control variable. To further control for the market trends, I also include the market-wide average of *LAMBDA* (*LAMBDA_MKT*) for each quarter. Finally, a dummy variable for a calendar year is included to control for the year fixed effect.

In the ETF-level regression, I include the daily percentage of intraday small trades averaged for a month (*ETF_PST*), and the average *ETF_LAMBDA* of all ETFs (*ETF_LAMBDA_MKT*) as additional control variables.

ETF diversification:

I construct two measures of the degree of diversification of an ETF. The first one is 0/1 dummy variable *(ETF_DVF1)*, which is 0 if an ETF is a sector ETF and 1 if an ETF is a diversified ETF based on its investment objective. To determine ETF investment objectives, I use both the “Morningstar Style box™” and ETF descriptions for investors.[20] The second measure of diversification is based on the factor analysis of three variables related to portfolio diversification: The number of stocks, the average of underlying stocks' weights, and the diversification score ranging from 0 to 3 based on investment objectives of ETFs.[21] *ETF_DVF2* is defined as high/low rank of this factor score, calculated each month.

As for firm-level observations during event quarters when new ETFs are introduced, I calculate *STK_DVF1* and *STK_DVF2*, dummy variables that measure the degree of diversification of ETFs in which a firm is newly included. *STK_DVF1* (*STK_DVF2*) is calculated with *ETF_DVF1* (*ETF_DVF2*) assigned to newly introduced ETFs as described above. In case a firm gets included in one ETF during an event quarter, *STK_DVF1* and *STK_DVF2* for the firm-quarter are simply *ETF_DVF1* and *ETF_DVF2* of the ETF holding the firm's shares. In the event of a firm getting included in multiple ETFs in one quarter, *ETF_DVF1* and *ETF_DVF2* of these ETFs are weighted by the number of shares held by these ETFs. For example, suppose two ETFs are introduced during a certain quarter and they both hold a firm's shares. 100 shares of the firm are held by one ETF with *ETF_DVF1*=0, and 200 shares of the firm are held by the other ETF with *ETF_DVF1*=1. Then the value-weighted diversification score for the firm-quarter is

[20] The Morningstar Style box is a 3 X 3 box with the growth level (value, blend, growth) along the vertical axis and size (large, mid, small) along the horizontal axis. Sample ETFs' positions on this nine-grid box are manually collected from http://corporate.morningstar.com. Furthermore, classifications for ETFs (e.g. large-cap growth ETF, financial industry ETF, etc.) provided by internet portals (Yahoo finance and Google finance) are collected. When the Morningstar Style box and ETF classifications do not coincide (e.g. funds that are not size-based are often categorized as mid-cap funds), I discretionally decide between the two from reading descriptions.

[21] The diversification score is 0 if an ETF is a sector ETF, 1 if an ETF duplicates size *and* growth based index, 1 if an ETF duplicates either a size-based index or a growth-based index, and 3 if an ETF duplicates a broad market index. The first diversification measure *ETF_DVF1* is 0 if this diversification score is 0 and 1 if it is between 1 and 3.

calculated to be 0.67. Because 0.67 is closer to 1 than to 0, *STK_DVF1* is defined to be 1 and the firm-quarter is considered to be an event set off by well-diversified ETFs. As well-diversified ETFs with *ETF_DVF1*=1 or *ETF_DVF2*=1 generally hold larger numbers of firms (and in many cases larger number of shares of each firm), the event study sample averages of these 0/1 dummy variables (*STK_DVF1* and *STK_DVF2*) are 0.83 and 0.87, respectively. That is, from 83% to 87% of the firm-quarters in the event study sample are categorized as events triggered by well-diversified ETFs.[22]

3.2. *Data Description*

I identify U.S. equity ETFs between 2002 and 2008, and obtain daily trading and quote data of these ETFs from the TAQ and daily market data from the CRSP. Figure 1 shows the number of ETFs used in the analysis. In the first sample quarter, a total of 63 ETFs are included in the sample, of which 32 ETFs are sector ETFs and 31 are diversified broader market index ETFs. In the last sample quarter, the number increases to 273, of which 133 are sector ETFs and 140 are diversified ETFs. Figure 2 shows the total market value of these sample ETFs. The total market value was approximately $99 billion in the first quarter of 2002, and grew five times by the end of 2008 to $448 billion. Although the numbers of ETFs are balanced between sector ETFs and diversified ETFs, the percentage of sector ETFs' market value is on average 13% of the total market value. I download these ETFs' holdings data from the Thomson S12 database and construct

[22] It is why I adopt diversification measures quarterly-ranked among ETFs, weight these rank dummies, and adjust the weighted dummies into 0/1 dummies for firm-level observations. If I rank diversification measures among the event firm-quarter observations in order to construct 0/1 dummies, the degree of diversification of newly introduced ETFs would be understated.

ETF-month observations. The final ETF-month sample includes 8,983 ETF-month observations of 294 ETFs.

All stocks actively traded between 2002 and 2008 are included in the firm-level sample. For these stocks, I obtain daily trading and quote data from the TAQ. I further require firm-quarter observations to have *SIZE, BTM, VOLUME, STDRET, ABS_ABNRET,* and *PRICE* calculated using the CRSP daily database and the COMPUSTAT quarterly database. In addition, *ANALYST* is calculated with the I/B/E/S detail data, and *INST_%* and *INST_N* are calculated with the I/B/E/S detail database and the Thomson Financial (S34) data. Missing values of *ANALYST*, *INST_%,* or *INST_N* are set to be zero. The final firm-quarter sample includes 152,151 firm-quarter observations of 8,420 firms. Using the holdings data of the sample ETFs, I calculate *ETF_%HLD* as the percentage of a firm's shares held by ETFs each quarter. The *ETF_%HLD* in figure 3 exhibits a similar pattern to figure 2. *ETF_%HLD* has grown from 0.8% in 2002 to 3.5% in 2008.[23] Diversified ETFs hold on average 10 times the number of firm shares that sector ETFs hold.

Table 1 presents size, growth, and the industry compositions of sample ETFs, and reports the average *ETF_LAMBDA* for each group.[24] The left panel shows that the sample is composed of 4,667 broad market index ETF-months and 4,316 sector ETF-months. Broad market index ETFs' average *ETF_LAMBDA* (0.022% of price) is 15% smaller

[23] In the first quarter of 2002, 54% of all actively traded firms were held by ETFs. In the last quarter of 2008, 80% of firms were held by ETFs.

[24] *LAMBDA* and *ETF_LAMBDA* are expressed in % of price. In eq. (4), φ represents the dealer cost component of bid-ask spreads, and λ represents the adverse selection cost component of bid-ask spreads, *LAMBDA*. Therefore, the implied bid-ask spreads calculated from eq. (4) is $(\varphi + \lambda)$, which is in % of price. The adverse selection component of bid-ask spreads in % of implied bid-ask spreads is calculated as $\lambda/(\varphi + \lambda)$.

than the sector ETFs' average *ETF_LAMBDA* (0.026% of price). The right panel summarizes the average *ETF_LAMBDA*s of ETFs that are excluded from the sample, either because they are not holding U.S. assets or because they are non-equity ETFs. The leveraged ETFs using derivatives based on domestic equities instead of holding stocks are categorized as non-equity ETFs, and are excluded from the sample. *ETF_LAMBDAs* are notably higher for these out-of-sample ETFs. For example, the average *ETF_LAMBDA* of non-U.S. ETFs is approximately twice higher than the average *ETF_LAMBDA* of the sample ETFs. The average *ETF_LAMBDA* of non-equity U.S. ETFs is approximately four times higher than that of the sample ETFs. The broad market index ETFs that are not in the sample exhibit an almost two times higher average *ETF_LAMBDA*, and the sector ETFs that are not in the sample exhibit an almost three times higher average *ETF_LAMBDA* than similar ETFs in the sample.[25] This comparison implies that information environments for domestic equities are richer than for international equities, commodities, or fixed assets. Also, higher return volatility provided by leveraged ETFs contributes to higher adverse selection of those ETFs. In summary, the sample selection procedure results in limiting the sample to relatively low adverse selection ETFs. [26]

---

[25] The difference between in-sample ETFs and out-of-sample ETFs is not as striking if I compare median values of *ETF_LAMBDA*s, but ETFs that are not in the sample largely exhibit higher median *ETF_LAMBDA*s.

[26] The reason why only U.S. equity ETFs are chosen is the availability of holdings data. This sample selection procedure results in the loss of a potentially interesting test setting of investors' migration to non-equity ETFs (e.g., leveraged ETFs, commodity ETFs, fixed-asset ETFs etc.) or international ETFs. However, limiting the sample to U.S. equity ETFs provides controlled setting to test the direct migration from individual stocks to ETFs holding those individual stocks due to private information. Also, the test of dynamics between lambdas of individual stocks and ETF lambdas is feasible by this limitation. If underlying stocks of international ETFs and the market data of these underlying stocks are easily acquirable,

Panel A of table 2 presents the descriptive statistics of the firm-quarter sample. The average *LAMBDA* is 0.159%, which is higher than the average *ETF_LAMBDA*s of most groups of ETFs in table 1. On average, 1.3% of firm shares are held by ETFs. *MFID_%HLD*, the percentage of a firm's shares held by index mutual funds, is on average 0.4%. , *MFAM_%HLD*, the percentage of firm shares held by actively managed mutual funds, is on average 9.6%, which is much higher than the averages of *ETF_%HLD* and *MFID_%HLD*. The average size of the sample firms is $2.2 billion and the average book-to-market ratio is 0.76. Panel A of table 3 shows that *LAMBDA* is significantly negatively correlated with *SIZE*, *VOLUME*, *PRICE*, *ANALYST*, *INST_%*, and *INST_N*.

Panel B of table 2 shows the descriptive statistics of changes in the variables in panel A, with the pre-period (post-period) set to be one quarter prior to (after) the event quarter. *LAMBDA* decreases by 0.001% on average over one quarter pre- and post-periods. *ETF_%HLD* increases by 0.4% on average. Surrounding these event quarters, the percentages of firm shares held by index or actively managed mutual funds do not change on average. The average institutional ownership increases both in percentage and in the number of institutions. Panel C of table 2 exhibits the descriptive statistics of *D_LAMBDA* and *D_ETF_%HLD* with the pre- and post-periods ranging from two to four quarters. Panel B of table 3 suggests that the change in *LAMBDA* is highly correlated with changes in *SIZE*, *BTM*, *VOLUME*, and *PRICE*. It is also noteworthy that the change in the percentage of institutional ownership exhibits the highest positive correlation with

they will constitute a sub-sample of higher adverse selection observations. I do not expect adding international ETF data changes the predictions and results of this paper.

D_ETF_%HLD. This supports the premise behind the first hypothesis that increased *ETF_%HLD* motivates uninformed investors to leave the market of individual stocks, and therefore the informed trade becomes more crowded.

Panel D of table 2 presents the descriptive statistics of the ETF-month sample. The average *ETF_LAMBDA* is 0.024% and is much smaller than the average *LAMBDA* (0.16%) of the firm-quarter sample in panel A of table 2. However, *ETF_LAMBDA* is on average not smaller than the value-weighted average of underlying stocks' *LAMBDA*s (*STK_LAMBDA*). The average size of the sample ETFs is $2 billion, and return volatility is notably lower for ETFs (0.25) than for individual stocks (17.12). Absolute abnormal return is also much lower for ETFs. Both the average percentage of institutional ownership and the number of institutional investors of ETFs are smaller than those of individual stocks. Panel C of table 3 shows that the two proxies for the degree of diversification of an ETF, *ETF_DVF1* and *ETF_DVF2*, are highly correlated with each other and exhibit similar correlations with other variables. *ETF_VOLUME* is highly correlated with *ETF_INST_%*, implying that although uninformed investors have stronger incentives to trade ETFs than informed investors, institutional investors also play an active role in trading ETFs. These institutional investors' trades could be informed ones based on their superior information about the market-wide factor or routine hedge activities against trading individual stocks or index futures.

4. Research Design and Test Results

In this section, I describe how I structure empirical tests. The tests examine the effect of an introduction of tradable baskets of stocks on the incentive of uninformed traders of underlying stocks of these baskets. I estimate regressions of the adverse selection component of bid-ask spreads on the percentage of a firm's shares held by ETFs and control variables

In section 4.1, I discuss and report results of levels regressions (table 4) and change regressions (tables 5 and 6) using the firm-quarter sample to test hypotheses 1-a and 1-b. In section 4.2, I discuss ETF-level regressions, for testing hypotheses 2-a and 2-b, and discuss results (table 7).

4.1. Firm-level tests

The first hypothesis predicts that the introduction of tradable baskets provides an incentive and an opportunity for uninformed investors to trade these baskets instead of individual stocks. As a result, the adverse selection cost for an individual stock is expected to increase.[27]

First, I estimate the following cross-sectional regression to investigate whether firms with a higher percentage of shares held by ETFs exhibit higher adverse selection costs. I estimate a pooled regression with two-way cluster robust standard errors (Gow, Ormazabal, and Taylor [2009]).

[27] Uninformed investors can switch from mutual funds to ETFs, not from individual stocks. Since these uninformed investors did not provide markets for individual stocks with liquidity before switching, the effect of this kind of switch on the adverse selection costs of individual stocks should be non-existent. Therefore, I assume the likelihood of this switch does not affect nor weakens my test results.

$$
\begin{aligned}
LAMBDA = {} & \alpha_{\mathrm{i}} + \beta_1 \cdot ETF\_\%HLD + \beta_2 \cdot SIZE + \beta_3 \cdot BTM + \beta_4 \cdot VOLUME \\
& + \beta_5 \cdot STDRET + \beta_6 \cdot ABS\_ABNRET + \beta_7 \cdot PRICE + \beta_8 \cdot ANALYST \\
& + \beta_9 \cdot INST\_\% + \beta_{10} \cdot INST\_N + \beta_{11} \cdot LAMBDA\_MKT + \beta_{12} \cdot PST \\
& + \beta_{13} \cdot YearDummy + \varepsilon_{\mathrm{it}}
\end{aligned} \tag{5}
$$

ETF_%HLD is the main test variable and defined to be the percentage of shares held by all U.S. equity ETFs in each quarter. I expect β_1 to be positive, consistent with the first hypothesis. The coefficient on *SIZE* is predicted to be negative, and the coefficient on *BTM* is predicted to be positive (Hedge and McDermott [2004b]). Trading volume is higher for large firms with liquid stocks. Therefore, I expect the association between *VOLUME* and *LAMBDA* to be negative. The return volatility and absolute abnormal returns indicate higher uncertainty during trading hours and inferior public disclosure. I expect both coefficients on *STDRET* and *ABS_ABNRET* to be positive. Prior studies suggest that stocks with higher prices are less liquid as some individual investors cannot afford the minimum trading units, leading to the prediction that β_7 will be positive.[28] However, higher prices are positively correlated with sizes of stocks, and therefore β_7 can be negative to the extent that price proxies for the firm size.

I expect β_8 to exhibit a negative sign as Brennan and Subrahmanyam [1995] finds a negative association between analyst following and adverse selection. Institutional ownership can positively affect adverse selection if it proxies for a large amount of

[28] Brennan and Subrahmanyam [1995] and Brennan and Subrahmanyam [1998] find that the price level is one of the important determinants of the lambda. Easley, O'Hara, and Saar [2001] finds that stock splits resulting in lower prices increase uninformed trades.

private information held by a few informed investors. On the other hand, if institutional ownership is higher only because institutional investors are numerous, the above Kyle [1985] model suggests that it leads to lower adverse selection due to competition among informed investors.[29] I include two proxies for the institutional ownership: *INST_%* and *INST_N*. The percentage of institutional ownership (*INST_%*) is predicted to be positively associated with *LAMBDA* given the number of institutional investors (*INST_N*) and other controls. Likewise, the number of institutional investors (*INST_N*) is predicted to be negatively associated with *LAMBDA* after I control for the level of *INST_%* and other variables.[30]

Panel A of table 4 reports the results of the regressions in eq. (5) with the firm-quarter sample. Column 1 presents a baseline regression on control variables. All coefficients exhibit predicted signs except on *ANALYST*. *PRICE* is negatively associated with *LAMBDA*. In the second column, I find the main test variable *ETF_%HLD*, the percentage of shares held by ETFs, exhibits a significant and positive association with *LAMBDA.* The result is consistent with predictions in hypothesis 1-a. The result is robust to controlling for the market average of *LAMBDA* (*LAMBDA_MKT*) and the percentage of small trades (*PST*) in the next two columns. In panel B of table 4, I include the percentage of shares held by mutual funds (*MFID_%HLD* and/or *MFAM_%HLD*) either as a control variable or as an alternative variable to *ETF_%HLD.* As predicted, the first

[29] Glosten and Harris [1988] suggests mixed associations between institutional or insider ownership and adverse selection. Van Ness, Van Ness, and Warr [2001] also provides mixed predictions on how institutional ownership affects adverse selection.
[30] The test results do not change qualitatively if I include either *INST_%* or *INST_N*. However, the main results are stronger if only *INST_N* is included and weaker if only *INST_%* is included than if both are controlled for. The explanatory power is highest when both are included.

three columns show that *MFID_%HLD* exhibits a similar effect on *LAMBDA* to *ETF_%HLD*. When both *ETF_%HLD* and *MFID_%HLD* are included in the regression, both are positively and significantly associated with *LAMBDA*. This result is robust when *MFAM_%HLD* is included. The result is also robust when I sum the *ETF_%HLD* and *MFID_%HLD*, and use it in the last column (*ID_%HLD*) as a proxy for passive investment opportunity.

As an additional test of hypothesis 1-a, I estimate the following regression using change variables with firm-quarters in which new ETFs are introduced. I estimate a pooled regression with two-way cluster robust standard errors (Gow et al. [2009]). I expect *D_LAMBDA*, the change in *LAMBDA*, to be positively associated with *D_ETF_%HLD*. I also expect other variables to exhibit the same signs as in the levels regression.

$$\begin{aligned} D\_LAMBDA = {} & \alpha_{\mathrm{i}} + \beta_1 \cdot D\_ETF\_\%HLD + \beta_2 \cdot D\_SIZE + \beta_3 \cdot D\_BTM \\ & + \beta_4 \cdot D\_VOLUME + \beta_5 \cdot D\_STDRET + \beta_6 \cdot D\_ABS\_ABNRET \\ & + \beta_7 \cdot D\_PRICE + \beta_8 \cdot D\_ANALYST + \beta_9 \cdot D\_INST\_\% \\ & + \beta_{10} \cdot D\_INST\_N + \beta_{11} \cdot D\_LAMBDA\_MKT + \beta_{12} \cdot D\_PST + \varepsilon_{\mathrm{it}} \end{aligned} \tag{6}$$

Table 5 reports the results of the regressions of the change in *LAMBDA* (*D_LAMBDA*) on the change in *ETF_%HLD* (*D_ETF_%HLD*) calculated over various lengths of pre- and post-periods. The first two columns of panel A of table 5 report the results of regressions using these change variables with one-quarter pre- and post-periods. *D_ETF_%HLD* over one quarter is positively associated with *D_LAMBDA*. As the period length increases, the association between *D_ETF_%HLD* and *D_LAMBDA* becomes more significant. As an additional 1% of a firm's shares are held by ETFs, *LAMBDA* increases by 0.003% over the next four quarters, which is approximately 2% of the average value of *LAMBDA* during the sample period. Note that when new ETFs are introduced, there is on average no change in *MFID_%HLD* and *MFAM_%HLD* (panel B of table 2). Nevertheless, *D_MFID_%HLD* affects *D_LAMBDA* while *D_MFAM_%HLD* exhibits a weak negative or no effect on *D_LAMBDA*, implying that the markets react differently to index funds and actively managed funds. In panel B of table 5, I use post-period variables instead of change variables for some of the control variables: *SIZE*, *BTM*, *ANALYST*, *INST_%*, and *INST_N*. These control variables represent relatively stable firm characteristics that are not expected to fluctuate over a few quarters. The result is robust to this alternative specification.

Hypothesis 1-b is based on the premise that a security-specific informed investor's incentive to become informed about funds is stronger for sector funds than for diversified funds. If a security-specific informed investor chooses to switch, being informed about a sector is less costly than being informed about the economy. This prediction implies that an increased *ETF_%HLD* (*D_ETF_%HLD)* by newly introduced diversified ETFs offers a stronger incentive for uninformed investors to choose these ETFs over an individual stock than when *ETF_%HLD* is increased by sector ETFs.

I estimate the change regression in eq. (7) using the interaction term of *D_ETF_%HLD* with either *STK_DVF1* or *STK_DVF2*. *STK_DVF1* and *STK_DVF2* are 0 if the firm-level event of being included in newly introduced ETFs is triggered by less diversified ETFs, and 1 if it is triggered by more diversified ETFs. I expect that the interaction term yields a positive coefficient β_3.

$$
\begin{aligned}
& D\_LAMBDA = \alpha_{\mathrm{i}} + \beta_1 \cdot STK\_DVF + \beta_2 \cdot D\_ETF\_\%HLD \\
& + \beta_3 \cdot D\_ETF\_\%HLD \times STK\_DVF + \beta_4 \cdot D\_SIZE + \beta_5 \cdot D\_BTM \\
& + \beta_6 \cdot D\_VOLUME + \beta_7 \cdot D\_STDRET + \beta_8 \cdot D\_ABS\_ABNRET \\
& + \beta_9 \cdot D\_PRICE + \beta_{10} \cdot D\_ANALYST + \beta_{11} \cdot D\_INST\_\% \\
& + \beta_{12} \cdot D\_INST\_N + \beta_{13} \cdot D\_LAMBDA\_MKT + \beta_{14} \cdot D\_PST \\
& + \beta_{15} \cdot D\_SIZE \times STK\_DFV + \beta_{16} \cdot D\_BTM \times STK\_DVF \qquad (7) \\
& + \beta_{17} \cdot D\_VOLUME \times STK\_DVF + \beta_{18} \cdot D\_STDRET \times STK\_DVF \\
& + \beta_{19} \cdot D\_ABS\_ABNRET \times STK\_DVF + \beta_{20} \cdot D\_PRICE \times STK\_DVF \\
& + \beta_{21} \cdot D\_ANALYST \times STK\_DVF + \beta_{22} \cdot D\_INST\_\% \times STK\_DVF \\
& + \beta_{23} \cdot D\_INST\_N \times STK\_DVF + \beta_{24} \cdot D\_LAMBDA\_MKT \times STK\_DVF \\
& + \beta_{25} \cdot D\_PST \times STK\_DVF + \varepsilon_{\mathrm{it}}
\end{aligned}
$$

Table 6 reports the results of the regression in eq. (7), and provides modest evidence supporting hypothesis 1-b. The first two columns demonstrate that, over one quarter pre- and post-periods, there is a weak significant difference between the introduction of diversified ETFs, and the introduction of sector ETFs in terms of the effect on *LAMBDA*. The regression results over two to three quarter pre- and post-periods, however, do not support this prediction. Over the longest pre- and post-periods, the introduction of well-diversified ETFs affects *D_LAMBDA* more strongly than less diversified ETFs do.

In summary, at the firm level, the availability of ETFs increases adverse selection of individual stocks as uninformed investors switch to trading ETFs. This effect is stronger if ETFs are more diversified.

4.2. ETF-level tests

Hypothesis 2-a suggests that ETFs that hold more diversified portfolios exhibit smaller adverse selection due to liquidity and diversification benefits, while hypothesis 2-b predicts that these ETFs exhibit larger adverse selection due to underlying stocks' increased adverse selection costs. Hypothesis 2-b contradicts hypothesis 2-a because of the endogenous nature of *STK_LAMBDA* introduced by hypothesis 1-a. That is, *STK_LAMBDA* increases as stocks are held by more diversified ETFs with lower *ETF_LAMBDA*, but at the same time *ETF_LAMBDA* increases in *STK_LAMBDA*. To address this endogeneity concern, I estimate the following simultaneous regression using the value-weighted average of stocks' idiosyncratic volatilities (*STK_IDIO*) as an instrumental variable. Stocks' idiosyncratic volatilities determine these stocks' adverse selection, but are expected to be exogenous to the adverse selection cost of an ETF that holds these stocks.[31]

The predicted value of *STK_LAMBDA* is calculated from the first-stage pooled regression in the following eq. (8). I estimate the second-stage pooled regression of *ETF_LAMBDA* on the predicted *STK_LAMBDA* as in the following eq. (9) with two-way cluster robust standard errors (Gow et al. [2009]).

[31] Idiosyncratic volatility is the residual return volatility after controlling for factors determining stock returns at the market-portfolio level (Ang, Hodrick, Xing, and Zhang [2006]). Spiegel and Wang [2005] finds that idiosyncratic volatility and illiquidity measures such as the Kyle [1995] lambda are positively associated. Although the study does not provide a clear explanation on why liquidity and idiosyncratic volatility are associated, it speculates that the association is the reason why both variables generate mixed views and empirical links to stock returns. I use the study by Spiegel and Wang [2005] as the support of using *STK_IDIO* as instrumental variable associated with *STK_LAMBDA*. Whether either liquidity or idiosyncratic volatility is priced or not is out of scope of this paper. I proceed with an assumption that idiosyncratic volatility is not priced, is diversified away at a portfolio level, and therefore does not affect portfolio-level variables such as *ETF_LAMBDA*. However, if idiosyncratic volatility is priced, my rational to use it as an exogenous variable to *ETF_LAMBDA* is weakened.

$$
\begin{aligned}
&STK\_LAMBDA = \alpha_{\text{it}} + \beta_1 \cdot STK\_IDIO + \beta_2 \cdot ETF\_SIZE \\
&+ \beta_3 \cdot ETF\_VOLUME + \beta_4 \cdot ETF\_STDRET \\
&+ \beta_5 \cdot ETF\_ABS\_ABNRET + \beta_6 \cdot ETF\_PRICE \\
&+ \beta_7 \cdot ETF\_INST\_\% + \beta_8 \cdot ETF\_INST\_N \\
&+ \beta_9 \cdot ETF\_LAMBDA\_MKT + \beta_{10} \cdot ETF\_PST + \varepsilon_{\text{it}}
\end{aligned} \tag{8}
$$

$$
\begin{aligned}
&ETF\_LAMBDA = \alpha_{\text{it}} + \beta_1 \cdot ETF\_DVF + \beta_2 \cdot STK\_LAMBDA \\
&+ \beta_3 \cdot STK\_LAMBDA \times ETF\_DVF + \beta_4 \cdot ETF\_SIZE + \beta_5 \cdot ETF\_VOLUME \\
&+ \beta_6 \cdot ETF\_STDRET + \beta_7 \cdot ETF\_ABS\_ABNRET + \beta_8 \cdot ETF\_PRICE \\
&+ \beta_9 \cdot ETF\_INST\_\% + \beta_{10} \cdot ETF\_INST\_N + \beta_{11} \cdot ETF\_LAMBDA\_MKT \\
&+ \beta_{12} \cdot ETF\_PST + \beta_{13} \cdot ETF\_SIZE \times ETF\_DVF \\
&+ \beta_{14} \cdot ETF\_VOLUME \times ETF\_DVF + \beta_{15} \cdot ETF\_STDRET \times ETF\_DVF \\
&+ \beta_{16} \cdot ETF\_ABS\_ABNRET \times ETF\_DVF + \beta_{17} \cdot ETF\_PRICE \times ETF\_DVF \\
&+ \beta_{18} \cdot ETF\_INST\_\% \times ETF\_DVF + \beta_{19} \cdot ETF\_INST\_N \times ETF\_DVF \\
&+ \beta_{20} \cdot ETF\_LAMBDA\_MKT \times ETF\_DVF \\
&+ \beta_{21} \cdot ETF\_PST \times ETF\_DVF + \varepsilon_{\text{it}}
\end{aligned} \tag{9}
$$

According to hypothesis 2-a, *ETF_LAMBDA* is lower for more diversified ETFs. Therefore, I expect *ETF_DVF* to exhibit a negative association with *ETF_LAMBDA*. The association between *STK_LAMBDA* and *ETF_LAMBDA* (β_2 or $\beta_2 + \beta_3$) can be either positive or negative. It will be positive to the extent that *STK_LAMBDA* represents the underlying assets' uncertainties which collectively determine *ETF_LAMBDA*. It will be negative to the extent that the outflow of uninformed investors from individual assets increases *STK_LAMBDA* (hypothesis 1-a).

Whether the association between *ETF_LAMBDA* and *STK_LAMBDA* differs between more diversified ETFs and less diversified ones is also an empirical question. If diversification of underlying stocks' uncertainties is the main driver of the association between *ETF_LAMBDA* and *STK_LAMBDA*, then β_3 on the interaction term is expected to exhibit a negative sign as the diversification benefit is stronger for more diversified portfolios (hypothesis 2-a). If the increase in *STK_LAMBDA* and the decrease in *ETF_LAMBDA* due to migration of uninformed trades is the main driver of the association between these two variables, then β_3 is expected to show a positive sign. Therefore, as a test of hypothesis 2-a (2-b), I predict that β_3 on the interaction term shows a negative (positive) sign.

Table 7 presents the results of the regressions in eq. (9). The first two columns present the second stage regression results with two diversification dummy variables that are calculated based on investment objectives (*ETF_DVF1*) and the factor analysis of three variables related to portfolio diversification (*ETF_DVF2*). Alternatively, the value weighted average of underlying stocks' *ETF_%HLD*s (*STK_ETF_%HLD*) is used as the endogenous variable. This usage of *STK_ETF_%HLD* instead of *STK_LAMBDA* is based on the first hypothesis that *ETF_%HLD* increases *LAMBDA*. Columns 3 and 4 present the second stage regression results with two diversification dummy variables on *STK_ETF_%HLD*.

As hypothesis 2-a predicts, *ETF_DVF* in the first column is negatively associated with *ETF_LAMBDA*. This implies that more diversified ETFs exhibit lower adverse selection because these ETFs become more liquid as more uninformed investors trade these ETFs. In contrast, the interaction term of *STK_LAMBDA* and *ETF_DVF* exhibits a positive coefficient supporting hypothesis 2-b that higher adverse selection costs of underlying stocks of more diversified ETFs lead to higher adverse selection costs of these ETFs. This result is largely robust to alternative measures of *ETF_DVF* and to using *STK_ETF_%HLD* as a proxy for *STK_LAMBDA.*

In summary, I attempt to separate the determinants of *ETF_LAMBDA* into the likelihood of informed trades, the diversification benefit, and the underlying stocks' uncertainties at the portfolio level. I find that smaller adverse selection costs of diversified ETFs attract more uninformed investors and this in turn makes stocks held by these diversified ETFs become even more illiquid.

4.3. Small trades

A key issue in this paper is to understand the effects when uninformed individual investors migrate to trading passive baskets of stocks. Therefore, I further investigate whether the availability of ETFs affects the percentage of intra-day small trades (*PST*) of individual stocks, which can be considered an alternative measure of the noise trades by uninformed investors. I define a small trade as one with a dollar value smaller than $5,000. The weakness of this measure as a direct measure of liquidity is the ad-hoc nature of the variable construction. For example, if a large institution has a capability of automating orders to frequently place small orders than $5,000 instead of one big order, then *PST* as a proxy for uninformed trades gets overstated.[32] Nevertheless, the analysis of

[32] Barclay and Warner [1993] and Chakravarty [2001] refer to the similar behavior of informed investors as "stealth-trading" and find that informed trades are concentrated on medium-sized trades (trades of 500-9,999 shares). Although these studies and my test use different set of samples, the percentage of medium-sized trades in their studies does not exceed the bottom 50%. Given that the average value *PST* is 0.72, most of small-trades in this paper may be categorized as medium-sized trades in their studies.

PST in the context of uninformed investors' migration seems to offer an interesting supplement to the main test in the paper.

First, I run the regression in eq. (5) with the dependent variable replaced by *PST* and report the results in panel A of table 8. Contrary to the prediction that *ETF_%HLD* decreases *PST* (to the extent that *PST* proxies for uninformed trades), the coefficient on *ETF_%HLD* is positive. *ID_%HLD*, the sum of *ETF_%HLD* and *MFID_%HLD*, also exhibit a positive association with *PST*. However, *MFID_%HLD* exhibits a negative and significant association with *PST*. Even though I am reassured by the negative sign of the coefficient on *MFID_%HLD*, the difference between *ETF_%HLD* and *MFID_%HLD* in their effects on *PST* is puzzling; this contrasts with the similarity in their influences on *LAMBDA*. I suspect that the uptrend of *PST* in the sample is too strong (from the average of 513/day in 2002 to 6,702/day in 2008) and biases the result, even though I include the market average (*PST_MKT*) as a control for the trend.

I run the change regression the eq. (6) with the dependent variable replaced by *D_PST*. Contrary to the levels regression result, the result in panel B of table 8 supports the prediction that the availability of ETFs decreases *PST*. In the first three columns, *D_ETF_%HLD* exhibits a negative and significant association with *D_PST*, implying that up to three quarters after the introduction of ETFs, *PST* decreases in *ETF_%HLD.*

I further run the IV regression similar to eq. (7) and eq. (8) with the dependent variable replaced by *D_PST*, and report the result on panel C of table 8. The interaction term of *D_ETF_%HLD* and *STK_DVF* exhibits a weak significant and negative association. The result confirms the prediction that more diversified ETFs have advantages in attracting individual investors. The comparison of table 6 and panel C of table 8 suggests that the impact of *D_ETF_%HLD* on *D_PST* is weaker but faster (in effect in one or two quarters) than the effect on *D_LAMBDA* (weakly in effect in a quarter but eventually in four quarters).

In summary, I find that the availability of index funds decreases small trades. The evidence is mixed in the equilibrium but clearer in the event study that presumably isolates the effect in a more efficient manner.

4.4. Self-selection problem

In this paper, the hypothesized positive association between *LAMBDA* and *ETF_%HLD* is attributed to the migration of uninformed investors, which, as a result, exacerbates adverse selection. However, one can suspect that ETFs tend to pick and hold stocks with higher *LAMBDA*s, as both the marginal benefit and the market demand can be larger for bundling stocks with higher *LAMBDA*s. If this is true, the positive association between *LAMBDA* and *ETF_%HLD* can be driven by the self-selection process rather than by the change in liquidity.

It is challenging to econometrically dismiss this alternative explanation, due to the difficulty of finding instrument variables uncorrelated with stock *LAMBDA.* Instead, I provide ETFs' investment objectives and descriptive statistics as evidence that ameliorates concerns about the self-selection issue. The ETFs in the sample are not actively managed by fund managers, but passively duplicate market indexes. Some indexes being duplicated are implicitly picking stocks with firm characteristics associated with high adverse selection (e.g., small-cap growth funds), but it is not a generalizable feature of all ETFs in the sample.

In figure 4, I present the comparison between the average *LAMBDA* of firms that are not included in any ETF in a given quarter (*ETF_%HLD*=0), and the average *LAMBDA* of firms held by ETFs (*ETF_%HLD*>0). Firms that are held by ETFs exhibit lower *LAMBDA*, supporting the argument that it is doubtful whether ETFs tend to hold high *LAMBDA* stocks. Figure 5 depicts the average pre-period *LAMBDA* of firms that are held by newly introduced ETFs during the event periods, and the average pre-period *LAMBDA* of the rest of the firms in the sample. During one quarter period prior to the event, firms that are included in new ETFs exhibit lower *LAMBDA*s than firms that are not picked by new ETFs. In summary, I argue that, although valid in principle, the self-selection concern should be minimal in the current data and tests presented in the paper.

5. Conclusion

This paper investigates whether the introduction of tradable baskets of stocks changes incentives of uninformed investors who find trading individual stocks without private information undesirable. The percentage of a firm's shares held by ETFs is used as a proxy for the uninformed investors' alternative trading opportunities. Using quarterly observations of all actively traded firms with required variables between 2002 and 2008, I find that the adverse selection cost of a stock increases in uninformed investors' opportunity to trade baskets of stocks. This positive association also holds in the regression of change variables calculated over one to four quarters prior and posterior to event quarters in which new ETFs are introduced. Uninformed investors are more likely to switch to trade a basket of stocks if the newly available ETF is more diversified, as they assess the likelihood of informed trades to be lower for more diversified ETFs than for less diversified ETFs. I find modest evidence on the stronger increase in the adverse selection costs for stocks held by newly introduced diversified ETFs than for stocks held by newly introduced sector ETFs. Overall, this increase in adverse selection due to the availability of ETFs is robust to various specifications of the tests.

As an additional test, I examine whether the portfolio-level adverse selection is lower for diversified ETFs than for sector ETFs or higher, as diversified ETFs are more likely to hold stocks with higher adverse selection costs. I find mixed evidence that potentially supports both of these predictions. ETFs' adverse selection costs decrease in diversification but the adverse selection costs of underlying stocks add up more significantly to the adverse selection costs of ETFs when these ETFs are more diversified.

In conclusion, the introduction of tradable baskets of stocks has an impact on uninformed investors' trading opportunities. Uninformed investors switch to trade baskets of stocks, and therefore the adverse selection cost of an individual stock increases in the availability of these tradable baskets. This result suggests that adverse selection is one of the costs of providing uninformed investors with lower risk trading devices. Even though this paper does not provide a comparison of the cost and the benefit of protecting uninformed investors, it is worthwhile to note that the new trading vehicles such as ETFs offered to uninformed investors can exacerbate the adverse selection of individual stocks, and can ultimately increase the adverse selection costs of ETFs.

References

ACKERT, L. AND Y. TIAN. "Arbitrage and Valuation in the Market for Standard and Poor's Depositary Receipts." *Financial Management* 29 (2000): 71-87.

ANG, A.; R. HODRICK; Y. XING; AND X. ZHANG. "The Cross-Section of Volatility and Expected Returns." *Journal of Finance* 61 (2006), 259-299.

ARMSTRONG, C.; J. CORE; D. TAYLOR; AND R. VERRECCHIA. "Asset Pricing in Imperfect Markets." Working paper, University of Pennsylvania, 2009.

ADMATI, A. AND P. PFLEIDERER. "A Theory of Intraday Patterns: Volume and Price Variability." *The Review of Financial Studies* 1 (1988): 3-40.

AMIHUD, Y. AND H. MENDELSON. "Dealership Market: Market-making with Inventory." *Journal of Financial Economics* 8 (1980): 31-53.

BARCLAY, M. AND J. WARNER. "Stealth trading and volatility: Which trades move prices?" *Journal of Financial Economics* 34 (1993), 281-305.

BERKMAN, H.; T. BRAILSFORD; AND A. FRINO. "A Note on Execution Costs for Stock Index Futures: Information versus Liquidity Effects." *Journal of Banking & Finance* 29 (2005): 565-577.

BRENNAN, M. AND A. SUBRAHMANYAM. "Investment Analysis and Price Formation in Securities Markets." *Journal of Financial Economics* 38 (1995): 361-381.

BRENNAN, M. AND A. SUBRAHMANYAM. "The Determinants of Average Trade Size." *Journal of Business* 71 (1998): 1-25.

CHAKRAVARTY, S. "Stealth-trading: Which traders' trades move stock prices?" *Journal of Financial Economics* 61 (2001): 289-307.

CHERKES M.; J. SAGI; AND R. STANTON. "A Liquidity-Based Theory of Closed-End Funds." *Review of Financial Studies* 22 (2009), 257-297.

CHRISTOFFERSEN, S.; D. KEIM; AND D. MUSTO. "Valuable Information and Costly Liquidity: Evidence from Individual Mutual Fund Trades." Working paper, University of Pennsylvania, 2006.

CLARKE, J. AND K. SHASTRI. "Adverse Selection Costs and Closed-End Funds." Working paper, Georgia Institute of Technology, 2001.

DICKSON, J.; J. SHOVEN; AND C. SIALM. “Tax Externalities of Equity Mutual Funds.” *National Tax Journal* 53 (2000): 607-628.

EASLEY, D.; R. ENGLE; M. O'HARA; AND L. Wu. “Time-Varying Arrival Rates of Informed and Uninformed Trades.” *Journal of Financial Econometrics* 6 (2008): 171-207.

EASLEY, D. AND M. O’HARA. “Price, Trade Size, and Information in Securities Markets.” *Journal of Financial Economics* 19 (1987): 69-90.

EASLEY, D.; M. O'HARA; AND G. SAAR. “How Stock Splits Affect Trading: A Microstructure Approach.” *Journal of Financial & Quantitative Analysis* 36 (2001): 25-51.

FOSTER, F. AND S. VISWANATHAN. “Variations in Trading Volume, Return Volatility, and Trading Costs - Evidence on Recent Price Formation Models.” *Journal of Finance* 48(1993): 187-211.

FRINO, A.; J. KRUK; AND A. LEPONE. “Transactions in Futures Markets: Informed or Uninformed?” *Journal of Futures Markets* 27 (2007): 1159-1174.

GAMMILL, J. AND A. PEROLD. “The Changing Character of Stock Market Liquidity.” *Journal of Portfolio Management* 15 (1989): 13-18.

GASTINEAU, G. “The Exchange-Traded Funds Manual.” Wiley and Sons, 2002.

GLOSTEN, L. AND L. HARRIS. “Estimating the Components of the Bid/Ask Spread.” *Journal of Financial Economics* 21 (1988): 123-142.

GLOSTEN, L. AND P. MILGROM. “Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders.” *Journal of Financial Economics* 14 (1985): 71-100.

GORTON, G., AND G. PENNACCHI. “Security Baskets and Index-Linked Securities.” *Journal of Business* 66 (1993): 1-27.

GOW, I.; G. ORMAZABAL: AND D. TAYLOR. “Correcting for Cross-Sectional and Time-Series Dependence in Accounting Research.” Working paper, Stanford University, 2009.

GUEDJ, I. AND J. HUANG. “Are ETFs Replacing Index Mutual Funds?” Working paper, University of Texas at Austin, 2009.

HASBROUCK, J. “Measuring the Information-Content of Stock Trades.” *Journal of Finance* 46 (1991): 179-207.

HASBROUCK, J. “Intraday Price Formation in U.S. Equity Index Markets.” *Journal of Finance* 58 (2003): 2375-2400.

HEGDE, S. AND J. MCDERMOTT. “Firm Characteristics as Cross-sectional Determinants of Adverse Selection.” *Journal of Business Finance & Accounting* 31 (2004): 1097-1124.

HEGDE, S. AND J. MCDERMOTT. “The Market Liquidity of DIAMONDS, Q's, and Their Underlying Stocks.” *Journal of Banking & Finance* 28 (2004): 1043-1067.

HENKER T. AND J. WANG. “On the importance of timing specifications in market microstructure research.” Journal of Financial Markets 9 (2006): 162-179.

HOLDEN, C. AND A. SUBRAHMANYAM. “Long-Lived Private Information and Imperfect Competition.” *Journal of Finance* 47 (1992): 247-270.

INVESTMENT COMPANY INSTITUTE, “2009 Investment Company Fact Book: A Review of Trends and Activity in the Investment Company Industry.” http://www.icifactbook.org/pdf/2009_factbook.pdf , 2009.

JEGADEESH, N. AND A. SUBRAHMANYAM. “Liquidity Effects of the Introduction of the S&P 500 Index Futures Contract on the Underlying Stocks.” *The Journal of Business* 66 (1993): 171-187.

KYLE, A. “Continuous Auctions and Insider Trading.” *Econometrica* 53 (1985): 1315-1335.

LEE C. AND M. READY. “Inferring Trade Direction from Intraday Data.” *Journal of Finance* 46 (1991): 733-746.

LEE, C.; A. SHLEIFER; AND R. THALER. “Investor Sentiment and the Closed-End Fund Puzzle.” *Journal of Finance* 46 (1991): 75-109.

LIN, J.; G. SANGER; AND G. BOOTH. “Trade Size and Components of the Bid-Ask Spread.” *The Review of Financial Studies* 8 (1995): 1153-1183.

MADHAVAN, A.; M. RICHARDSON; AND M. ROOMANS. “Why Do Security Prices Change? A Transaction-Level Analysis of NYSE Stocks.” *The Review of Financial Studies* 10 (1997): 1035-1064.

MADHAVAN, A. AND S. SMIDT. “An Analysis of Changes in Specialist Inventories and Quotations.” *Journal of Finance* 48 (1993): 1595-1628.

MARKOWITZ, H. “Portfolio Selection.” *Journal of Finance* 7 (1952), 77-91.

NEAL, R. AND S. WHEATLEY. “Adverse Selection and Bid–Ask Spreads: Evidence from Closed-End Funds.” *Journal of Financial Markets* 1 (1998): 121-149.

PONTIFF, J. “Costly Arbitrage: Evidence from Closed-End Funds.” *The Quarterly Journal of Economics* 111 (1996): 1135-1151.

PONTIFF, J. “Costly Arbitrage and the Myth of Idiosyncratic Risk.” *Journal of Accounting and Economics* 42 (2006): 35-52.

ROLL, R.; E. SCHWARTZ; AND A. SUBRAHMANYAM. “Liquidity and the Law of One Price: The Case of the Futures-Cash Basis.” *Journal of Finance* 62 (2007): 2201-2234.

SPIEGEL, M. AND A. SUBRAHMANYAM. “Informed Speculation and Hedging in a Noncompetitive Securities Market.” *The Review of Financial Studies* 5 (1992): 307-329.

SPIEGEL, M. AND X. WANG. “Cross-sectional Variation in Stock Returns: Liquidity and Idiosyncratic Risk.” Working paper, Yale University, 2005.

STOLL, H. “Friction.” *Journal of Finance* 55 (2000): 1479-1514.

SUBRAHMANYAM, A. “A Theory of Trading in Stock Index Futures.” *The Review of Financial Studies* 4 (1991): 17-51.

VAN NESS, B.; R. VAN NESS; AND R. WARR. “How well do adverse selection components measure adverse selection?” *Financial Management* 30 (2001): 77-98.

VERRECCHIA, R. “Essays on disclosure.” *Journal of Accounting & Economics* 32 (2001): 97-180.

ZINGALES, L. “The Future of Securities Regulation.” *Journal of Accounting Research* 47 (2009): 391-425.

Appendixes

A.1 Imperfect competition model

The value of a firm disclosed by the firm is denoted by $\widetilde{y} = \widetilde{u} + \widetilde{\eta}$, where $\widetilde{u}\,\tilde{}\,N(m, \frac{1}{h})$ and $\widetilde{\eta}\,\tilde{}\,N(0, \frac{1}{n})$. h represents the precision assessed by the markets on the ex-ante value of an asset $\widetilde{u}$, and n is the precision of the information content of the disclosure $\widetilde{y}$ (Verrecchia [2001]). The informed investor knows $\widetilde{\eta} = \eta$ whereas the market maker only knows $\widetilde{y} = y$. The informed investor determines his or her demand d based on knowledge of $\widetilde{u}$ and the demand order gets batched with demand orders generated from random shocks $\widetilde{x}$. The total demand is then $\widetilde{D} = d + \widetilde{x}$, where $\widetilde{x}\,\tilde{}\,N(0, \frac{1}{t})$. The market maker executes this combined order at a single price p that is equivalent to the market maker's expectation of the firm value conditional on the disclosure and the total demand $E\left(\widetilde{u}|\widetilde{y}, \widetilde{D}\right)$. Before submitting the informed order d, the informed investor expects how the market will come up with this price p based on $\widetilde{y}$ and $\widetilde{D}$. The informed investor conjectures the following eq.(A1.1), where β represents the price reaction to surprise in disclosure and λ represents information asymmetry.

$$p = m + \beta(y - m) + \lambda D \tag{A1.1}$$

Then, the informed investor chooses the order amount d that maximizes his or her expected profit.

$$\underset{d}{Max}\left[dE\left(\widetilde{u} - \widetilde{p}|\widetilde{u} = u, \widetilde{y} = y\right)\right]$$

$$= \underset{d}{Max}\left[dE(\widetilde{u} - (m + \beta(y - m) + \lambda(d + \widetilde{x}))\,|\widetilde{u} = u, \widetilde{y} = y)\right]$$

$$= \underset{d}{Max}\left[d\left(u - m - \beta y + \beta m - \lambda d\right)\right]$$

The first order condition with regard to the investor's demand leads to the following optimal demand.

$$d^* = \frac{1}{2\lambda}\left(u - m - \beta\left(y - m\right)\right) \tag{A1.2}$$

The market efficiency condition suggests that an asset's price is the market's expection of the asset's value, conditional upon all information such that $p = E\left(\widetilde{u}|\widetilde{y}, \widetilde{D}\right)$.

Combined with eq.(A1.2), the market efficiency condition yields:

$$p = E\left(\widetilde{u}|\widetilde{y}, \widetilde{D}\right) = m + \left(\frac{4\lambda^2 n + t\beta}{4\lambda^2 n + t + 4\lambda^2 h}\right)(y - m) + \left(\frac{2\lambda t}{4\lambda^2 n + t + 4\lambda^2 h}\right) D \tag{A1.3}$$

From eq.(A1.1) and eq.(A1.3), β and λ are calculated as $\beta = \dfrac{n}{h+n}$ and $\lambda = \dfrac{1}{2}\sqrt{\dfrac{t}{h+n}}$.

Therefore, the equilibrium price is solved as in the following eq.(A1.4).

$$\widetilde{p} = m + \frac{n}{h+n}(\widetilde{y} - m) + \frac{1}{2}\sqrt{\frac{t}{h+n}}\widetilde{D} \tag{A1.4}$$

When there are N identical informed investors instead of one informed investor, each informed chooses d to maximize his or her expected profit, expecting other informed investors would choose $\widehat{d}$.

$$\underset{d}{Max}\left[dE\left(\widetilde{u} - \widetilde{p}|\widetilde{u} = u, \widetilde{y} = y\right)\right]$$

$$= \underset{d}{Max}\left[dE(\widetilde{u} - (m + \beta(y - m) + \lambda(Nd + M\widetilde{x}))\left|\widetilde{u} = u, \widetilde{y} = y\right)\right]$$

$$= \underset{d}{Max}\left[d\left(u - m - \beta y + \beta m - \lambda\left(d + (N-1)\widehat{d} + x\right)\right)\right]$$

The first order condition with regard to the investor's demand leads to the following optimal demand.

$d^* = \dfrac{1}{2\lambda}\left(u - m - \beta\left(y - m\right) - \lambda(N-1)\widehat{d}\right)$

Since informed investors are identical, $d^* = \widehat{d}$. Therefore,

$$d^* = \frac{1}{\left(N+1\right)\lambda}\left(u - m - \beta\left(y - m\right)\right) \tag{A1.5}$$

Combined with eq.(A1.1) and eq.(A1.5), the market efficiency condition $p = E\left(\widetilde{u}|\widetilde{y}, \widetilde{D}\right)$ yields:

$\beta = \dfrac{n}{h+n}$ and $\lambda = \dfrac{\sqrt{Nt}}{(1+N)}\sqrt{\dfrac{1}{h+n}}$.

Therefore, the equilibrium price with N informed investors is solved as in the following eq.(A1.6).

$$\widetilde{p} = m + \frac{n}{h+n}(\widetilde{y} - m) + \frac{\sqrt{Nt}}{(1+N)}\sqrt{\frac{1}{h+n}}\widetilde{D} \tag{A1.6}$$

A.2. Portfolio lambda

Let V the variance-covariance matrix of a portfolio. Let e denote a vector of ones such that $e = \begin{bmatrix} 1 & 1 & .... & 1 \end{bmatrix}^T$. The vector of optimal weights w^* that minimizes the portfolio variance solves the following problem.

$w^* = \arg\min\limits_{w}\left(w^T V w\right)$,
subject to $w^T e = e$.

With the Lagrangian multiplier θ, the problem becomes:

$$\underset{w}{Min}\left[g\left(w,\theta\right)\right]=\underset{w}{Min}\left[w^TVw+\theta\left(1-w^Te\right)\right]$$

The first order condition with regard to the vecotor of weights yields:

$$\frac{\partial g}{\partial w}=2Vw-\theta e=0\Leftrightarrow w=\frac{\theta}{2}V^{-1}e$$

From the constraint that the sum of weights is one, we get

$$w^Te=\left(\frac{\theta}{2}V^{-1}e\right)^T e=\frac{\theta}{2}e^TV^{-1}e=1\Leftrightarrow\theta=\frac{2}{e^TV^{-1}e}$$

Combined with the first order condition, the vector of optimal weights is solved as the following eq.(A2.1).

$$w^*=\frac{1}{e^TV^{-1}e}V^{-1}e \tag{A2.1}$$

Therefore, the overall uncertainty of the optimaly weighted portfolio is cacluated as the following eq.(A2.2).

$$w^{*T}Vw^*=\frac{1}{\left(e^TV^{-1}e\right)^2}e^TV^{-1}VV^{-1}e=\frac{1}{e^TV^{-1}e} \tag{A2.2}$$

, which is equivalant to $\frac{1}{h+n}$ of a single asset. Therefore, portfolio-level lambda λ_{pf} is calculated as

$$\lambda_{pf}=\frac{\sqrt{N_{pf}t_{pf}}}{(1+N_{pf})}\sqrt{\frac{1}{e^TV^{-1}e}} \tag{A2.3}$$

,where N_{pf} is the number of informed traders trading the portfolio, t_{pf} is the precision of noise trading.

A.3 The number of assets in the portfolio

Let V_J denote the optimally weighted portfolio's uncertainty defined as in eq.(A2.2) constructed with J assets.

Assume a portfolio with $(J+1)$ assets is constructed by adding one more asset to the portfolio with J assets. V_{J+1} is written as the following.[1]

$$V_{J+1} = \begin{bmatrix} V_J & b \\ b^T & v_{J+1} \end{bmatrix}, \text{ where } b = \begin{bmatrix} Cov(\widetilde{u}_1, \widetilde{u}_{J+1}|y) & .... & Cov(\widetilde{u}_J, \widetilde{u}_{J+1}|y) \end{bmatrix}^T$$

and $v_{J+1} = Var(\widetilde{u}_{J+1}|y)$,where y is a vector of disclosed values of $J+1$ assets.

Let k denote the Schur complement $k = v_{J+1} - b^T V_J^{-1} b$. Since both V_J and V_{J+1} are positive-definite, k is also positive. Using k, the inverse of V_{J+1} is written as the following.

$$V_{J+1}^{-1} = \begin{bmatrix} V_J & b \\ b^T & v_{J+1} \end{bmatrix}^{-1} = \begin{bmatrix} V_J^{-1} + \frac{1}{k} V_J^{-1} b b^T \left(V_J^{-1}\right)^T & -\frac{1}{k} V_J^{-1} b \\ -\frac{1}{k} b^T \left(V_J^{-1}\right)^T & \frac{1}{k} \end{bmatrix} \tag{A3.1}$$

Let $P(V_J)$ denote the precision of the optimally weighted portfolio, which is equivalent to the inverse of V_J. Then $P(V) = e^T V_J^{-1} e$. Let $IP_{(J,J+1)}$ denote the incremental portfolio precision when an asset is added to the portfolio with J assets.

$$IP_{(J,J+1)} = P(V_{J+1}) - P(V_J) = e^T V_{J+1}^{-1} e - e^T V_J^{-1} e \tag{A3.2}$$

[1] Let X_N denote a portfolio-level variable X when the number of asset of the portfolio is N.

From Eq (A3.1) and Eq.(A3.2),

$$IP_{(J,J+1)} = \frac{1}{k} e^T V_J^{-1} b b^T \left(V_J^{-1}\right)^T e - \frac{2}{k} e^T V_J^{-1} b + \frac{1}{k} = \frac{\left(e^T V_J^{-1} b - 1\right)^2}{v_{J+1} - b^T V_J^{-1} b} \geq 0$$

Therefore, as the number of assets in a portfolio increases, the uncertainty about the portfolio value decreases.

A.4 Increase in uncertainly about an individual asset

$IP_{(1,2)} = P(V_2) - P(V_1)$,

where $V_1 = v_1$, $V_2 = \begin{bmatrix} v_1 & b \\ b^T & v_2 \end{bmatrix}$ and $-\sqrt{v_1 v_2} < b < \sqrt{v_1 v_2}$.

Then, $\dfrac{\partial}{\partial v_2} IP_{(1,2)} = \dfrac{\partial}{\partial v_2} \left(\dfrac{b^2 - 2bv_1 + v_1^2}{v_1^2 v_2 - b^2 v_1} \right) \leq 0.$

Therefore, as the uncertainty about each asset's value increases, the uncertainty about the portfolio value increases.

A.5 Covariances among asset values

$V_1 = v_1$ and $V_2 = \begin{bmatrix} v_1 & b \\ b^T & v_2 \end{bmatrix}$,where $-\sqrt{v_1 v_2} < b < \sqrt{v_1 v_2}$.

The incremental precision is $IP_{(1,2)} = \dfrac{\left(\dfrac{b}{v_1} - 1\right)^2}{v_2 - \dfrac{b^2}{v_1}}$.

$$\frac{\partial}{\partial b} IP_{(1,2)} = \frac{\partial}{\partial b} \frac{\left(\dfrac{b}{v_1} - 1\right)^2}{v_2 - \dfrac{b^2}{v_1}} = \frac{-\dfrac{2}{v_1^2}(b - v_1)(b - v_2)}{\left(v_2 - \dfrac{b^2}{v_1}\right)^2}$$

$$\Longleftrightarrow \frac{\partial}{\partial b} IP_{(1,2)} > 0 \text{ if } \min(v_1, v_2) < b < \max(v_1, v_2) \text{ and } \frac{\partial}{\partial b} IP_{(1,2)} < 0\text{,otherwise.}$$

Since $\min(v_1, v_2) < \sqrt{v_1 v_2} < \max(v_1, v_2)$,

$\frac{\partial}{\partial b} IP_{(1,2)} \leq 0$, if $-\sqrt{v_1 v_2} < b \leq \min(v_1, v_2)$ and

$\frac{\partial}{\partial b} IP_{(1,2)} > 0$, if $\min(v_1, v_2) < b < \sqrt{v_1 v_2}$

Therefore, the uncertainty about the portfolio value is small when covariances among assets are negative or very small. The result implies that the more diversified a portfolio, the larger the uncertainty of the portfolio value. Generalization of A.5 for $J + 1$ asset portfolio is as follows.

The uncertainty about $J+1$ asset portfolio value is denoted by $V_{J+1} = \begin{bmatrix} V_J & b \\ b^T & v_{J+1} \end{bmatrix}$, where $b = \begin{bmatrix} b_1 & b_2 & .... & b_J \end{bmatrix}^T$.

From Eq.(A3.2),

$$\frac{\partial}{\partial b_j} IP_{(J,J+1)} = 2 \frac{\left(e^T V_J^{-1} b - 1\right)}{\left(v_{J+1} - b^T V_J^{-1} b\right)} \left[\left[e^T V_J^{-1}\right]_j + 2 \left(e^T V_J^{-1} b - 1\right) \left[b^T V_J^{-1}\right]_j \right].$$

Let $\beta = e^T b$.Then, $\frac{\partial}{\partial \beta} IP_{(J,J+1)} = \sum_{j=1}^{J} \frac{\partial}{\partial b_j} IP_{(J,J+1)} = \frac{4}{k} (t - 1) \left(t^2 - t + \frac{p}{2}\right)$,

$$\text{where} \begin{pmatrix} t = e^T V_J^{-1} b \\ k = v_{J+1} - b^T V_J^{-1} b > 0 \\ p = e^T V_J^{-1} e > 0 \end{pmatrix}.$$

Therefore,

(a) When $e^T V_J^{-1} e \geq \frac{1}{2}$:

$$\frac{\partial IP_{(1,2)}}{\partial \beta} \leq 0, \text{ if } e^T V_J^{-1} b \leq 1$$

$$\frac{\partial IP_{(1,2)}}{\partial \beta} > 0, \text{ if } e^T V_J^{-1} b > 1$$

(b) When $0 < e^T V_J^{-1} e < \frac{1}{2}$:

$$\frac{\partial IP_{(1,2)}}{\partial \beta} \leq 0,$$

$$\text{if } \frac{1}{2} + \frac{1}{2}\sqrt{1 - 2\left(e^T V_J^{-1} e\right)} \leq e^T V_J^{-1} b \leq 1 \text{ or } e^T V_J^{-1} b \leq \frac{1}{2} - \frac{1}{2}\sqrt{1 - 2\left(e^T V_J^{-1} e\right)}$$

$$\frac{\partial IP_{(1,2)}}{\partial \beta} > 0,$$

$$\text{if } e^T V_J^{-1} b > 1 \text{ or } \frac{1}{2} - \frac{1}{2}\sqrt{1 - 2\left(e^T V_J^{-1} e\right)} < e^T V_J^{-1} b < \frac{1}{2} + \frac{1}{2}\sqrt{1 - 2\left(e^T V_J^{-1} e\right)}.$$

A.6 Dispersion of elements of b on incremental precision

In A.5, the magnitude of covariances is assumed to represent the extent to which a portfolio is diversified. As an additional measure of diversification, I examine the dispersion of covariances among asset values.

Let $V_2 = \begin{bmatrix} v & b \\ b & v \end{bmatrix}$, where $v_1 > v_2$

and let $V_3 = \begin{bmatrix} v & b & b+\varepsilon \\ b & v & b-\varepsilon \\ b+\varepsilon & b-\varepsilon & v \end{bmatrix}$,where $\varepsilon \geq 0$, $\begin{bmatrix} b+\varepsilon \\ b-\varepsilon \end{bmatrix} = be + \varepsilon\pi$ and $\pi = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

The incremental precision is $IP_{(2,3)} = P(V_3) - P(V_2)$

$$= e^T V_3^{-1} e - e^T V_2^{-1} e = \frac{\left(\dfrac{2b}{v+b} - 1\right)^2}{v - \dfrac{2b^2}{v+b} - \dfrac{2\varepsilon^2}{v-b}}$$

Therefore, $\dfrac{\partial IP_{(2,3)}}{\partial \varepsilon} = \dfrac{\partial}{\partial \varepsilon}\left(\dfrac{\left(\dfrac{2b}{v+b} - 1\right)^2}{v - \dfrac{2b^2}{v+b} - \dfrac{2\varepsilon^2}{v-b}}\right) = \dfrac{4\varepsilon\left(\dfrac{2b}{v+b} - 1\right)^2}{\left(v - \dfrac{2b^2}{v+b} - \dfrac{2\varepsilon^2}{v-b}\right)^2} > 0$

The result also implies that the more diversified a portfolio, the larger the uncertainty of the portfolio value.

Table 1: LAMBDAs of ETFs in the sample and ETFs excluded from the sample

		SAMPLE						NOT IN SAMPLE					
		N	Mean	Std Dev	Min	Median	Max	N	Mean	Std Dev	Min	Median	Max
BROAD MARKET		**4667**	**0.022**	**0.049**	**0.000**	**0.011**	**1.968**	**4588**	**0.050**	**0.122**	**0.000**	**0.016**	**3.508**
BLEND		**1744**	**0.023**	**0.069**	**0.000**	**0.010**	**1.968**	**2193**	**0.050**	**0.134**	**0.000**	**0.014**	**3.508**
	Large-cap	841	0.017	0.028	0.000	0.008	0.322	1511	0.041	0.125	0.000	0.013	3.508
	Mid-cap	400	0.023	0.102	0.000	0.008	1.968	317	0.063	0.155	0.000	0.023	2.179
	Small-cap	467	0.030	0.075	0.000	0.014	1.318	64	0.157	0.237	0.000	0.079	1.341
	Multi-cap	36	0.083	0.138	0.001	0.034	0.661	301	0.058	0.108	0.000	0.020	1.054
GROWTH		**1321**	**0.023**	**0.036**	**0.000**	**0.013**	**0.446**	**681**	**0.048**	**0.114**	**0.000**	**0.019**	**2.179**
	Large-cap	479	0.020	0.039	0.000	0.010	0.327	382	0.041	0.080	0.000	0.017	1.072
	Mid-cap	447	0.027	0.037	0.000	0.016	0.446	49	0.075	0.035	0.018	0.069	0.154
	Small-cap	336	0.022	0.031	0.000	0.013	0.278	68	0.029	0.037	0.000	0.014	0.201
	Multi-cap	59	0.029	0.034	0.000	0.014	0.149	182	0.061	0.184	0.001	0.015	2.179
VALUE		**1602**	**0.019**	**0.029**	**0.000**	**0.011**	**0.398**	**1714**	**0.051**	**0.108**	**0.000**	**0.017**	**1.282**
	Large-cap	798	0.017	0.025	0.000	0.011	0.334	1408	0.047	0.102	0.000	0.016	1.282
	Mid-cap	411	0.020	0.036	0.000	0.010	0.398	220	0.073	0.142	0.001	0.030	1.212
	Small-cap	350	0.021	0.026	0.000	0.013	0.240	40	0.090	0.141	0.000	0.042	0.745
	Multi-cap	43	0.020	0.025	0.000	0.014	0.161	46	0.014	0.015	0.001	0.009	0.078
large-cap		**2118**	**0.018**	**0.030**	**0.000**	**0.010**	**0.334**	**3301**	**0.044**	**0.111**	**0.000**	**0.015**	**3.508**
mid-cap		**1258**	**0.023**	**0.065**	**0.000**	**0.012**	**1.968**	**586**	**0.068**	**0.144**	**0.000**	**0.028**	**2.179**
small-cap		**1153**	**0.025**	**0.053**	**0.000**	**0.018**	**1.318**	**529**	**0.091**	**0.170**	**0.000**	**0.015**	**1.341**
multi-cap		**138**	**0.040**	**0.079**	**0.000**	**0.013**	**0.661**	**172**	**0.055**	**0.136**	**0.000**	**0.030**	**2.179**
		N	Mean	Std Dev	Min	Median	Max	N	Mean	Std Dev	Min	Median	Max
SECTOR		**4316**	**0.026**	**0.066**	**0.000**	**0.012**	**1.968**	**1602**	**0.074**	**0.173**	**0.000**	**0.015**	**2.169**
Commodity		3	0.065	0.060	0.013	0.052	0.130	5	0.976	0.502	0.583	0.659	1.558
Construction		35	0.033	0.021	0.007	0.029	0.084						
Consumer discretionary		224	0.022	0.027	0.000	0.012	0.178	78	0.050	0.182	0.000	0.007	1.369
Consumer goods		89	0.025	0.044	0.000	0.007	0.230						
Consumer service		88	0.032	0.052	0.000	0.012	0.332	42	0.121	0.056	0.019	0.119	0.262
Consumer staples		134	0.013	0.021	0.000	0.007	0.171	1	0.000	.	0.000	0.000	0.000

Defense	20	0.078	0.123	0.000	0.015	0.398	13	0.222	0.104	0.103	0.192	0.417
Energy	411	0.025	0.032	0.000	0.015	0.319	98	0.076	0.224	0.001	0.008	1.704
Financial	507	0.027	0.090	0.000	0.016	1.968	214	0.186	0.227	0.000	0.107	1.395
Food	20	0.015	0.014	0.000	0.015	0.050						
Healthcare	516	0.013	0.017	0.000	0.008	0.172	190	0.031	0.106	0.000	0.007	1.212
Home builders	37	0.048	0.090	0.007	0.015	0.398						
Industrial	239	0.024	0.043	0.000	0.011	0.327	30	0.157	0.407	0.004	0.034	2.169
MISC	11	0.167	0.099	0.034	0.169	0.332	1	0.000	.	0.000	0.000	0.000
Materials	25	0.163	0.387	0.000	0.064	1.968	22	0.027	0.034	0.001	0.014	0.121
Natural resources	344	0.021	0.024	0.000	0.012	0.162	1	0.001	.	0.001	0.001	0.001
Real estate	347	0.027	0.049	0.000	0.012	0.446	181	0.061	0.112	0.003	0.021	1.072
Retail	29	0.018	0.011	0.001	0.012	0.051						
Tech	748	0.033	0.083	0.000	0.014	1.374	472	0.061	0.146	0.000	0.014	1.453
Telecom	209	0.026	0.084	0.000	0.011	1.151	170	0.017	0.021	0.000	0.010	0.121
Transportation	44	0.014	0.013	0.000	0.012	0.054						
Utilities	236	0.015	0.027	0.000	0.008	0.341	84	0.016	0.022	0.000	0.008	0.107
US	**8983**	**0.024**	**0.058**	**0.000**	**0.011**	**1.968**	**1813**	**0.080**	**0.174**	**0.000**	**0.019**	**2.179**
NON-US							**4377**	**0.046**	**0.118**	**0.000**	**0.015**	**3.508**

This table provides the average **LAMBDA** of ETFs grouped by size, growth, or industry. **LAMBDA** (in % of price) is the adverse selection component (λ) of bid-ask spreads estimated per month with the method by Madhavan et al. [1997], modified by Armstrong et al. [2009]. Observations with negative **LAMBDA**s or with ones larger than 100% are excluded from the calculation.

Table 2: Descriptive Statistics

Panel A: Firm-quarter observations of levels variables

Variable	N	Mean	Std Dev	Min	Median	Max
LAMBDA	152,151	0.159	0.331	0.000	0.044	4.093
ETF_%HLD	152,151	0.013	0.017	0.000	0.007	0.284
MFID_%HLD	152,151	0.004	0.007	0.000	0.000	0.028
MFAM_%HLD	152,151	0.096	0.095	0.000	0.072	0.398
SIZE	152,151	2,195.80	6,472.36	1.77	305.56	60,667.13
BTM	152,151	0.761	0.372	0.089	0.762	2.694
VOLUME	152,151	-6.052	1.717	-12.365	-5.636	-3.042
STDRET	152,151	17.124	3.321	0.811	16.848	34.157
ABS_ABNRET	152,151	0.168	0.177	0.000	0.115	1.960
PRICE	152,151	20.17	18.68	0.19	15.15	121.38
ANALYST	152,151	4.840	6.050	0.000	3.000	29.000
INST_%	152,151	0.449	0.335	0.000	0.439	1.000
INST_N	152,151	94.23	126.92	0.00	51.00	842.00
PST	152,151	0.723	0.263	0.000	0.808	1.000
LAMBDA_MKT	152,151	0.165	0.102	0.078	0.108	0.445

LAMBDA (in %) is the adverse selection component (λ) of bid-ask spreads estimated per month with the method by Madhavan et al. [1997], modified by Armstrong et al. [2009]. Observations with negative values or larger than 100(%) of **LAMBDA** are excluded from the sample. **ETF_%HLD** is the percentage of shares held by any of the U.S. equity ETFs, calculated with Thomson holdings data (S12). The holding data are reported each month or quarter. **MFID_%HLD** is the percentage of shares held by U.S. equity index mutual funds, calculated with Thomson holdings data (S12). The holding data are reported each month or quarter. **MFAM_%HLD** is the percentage of shares held by actively managed U.S. equity mutual funds. All non-ETF, non-index, U.S. equity mutual funds in Thomson holdings data (S12) are categorized as actively managed U.S. equity mutual funds. The holding data are reported each month or quarter. **ETF_%HLD, MFID_%HLD**, and **MFAM_%HLD** are calculated daily and averaged by quarter. **SIZE** is the log of one plus market value (CRSP *prc* * *shrout*). Tabulated are unlogged values of **SIZE** in millions. **BTM** is the book-to-market assets ratio calculated as (Compustat *ATQ*/(*ATQ* -Compustat *CEQQ* + Market Value))) on the quarter end date. **VOLUME** is the turnover calculated as the log of one plus the quarterly average of daily CRSP trading volume divided by one plus the number of shares outstanding. **STDRET** is the annualized time-series standard deviation of daily returns (CRSP *ret*), calculated each quarter. **ABS_ABNRET** is the absolute abnormal cumulative returns in each quarter. **PRICE** is the log of one plus the quarterly average of daily CRSP prices. Tabulated are the unlogged values of **PRICE**. **ANALYST** is the log of one plus the quarterly average number of analysts following counted from I/B/E/S. Tabulated are the unlogged values of **ANALYST. INST_%** is the quarterly % of institutional ownership from Thomson Financial Spectrum. **INST_N** is the log of one plus the quarterly number of institutional owners from Thomson Financial Spectrum. Tabulated are the unlogged values of **INST_N. PST** is the % of small trades (defined as trades smaller than $5,000), calculated with the TAQ intraday trade dataset. **PST** is calculated daily and averaged by quarter. **LAMBDA_MKT** is the average **LAMBDA** of all stocks in the sample. All variables are winsorized at 1%.

Panel B: Firm-quarter observations of change variables over one-quarter pre- & post-periods

No. QTRs: 1

Variable	N	Mean	Std Dev	Min	Median	Max
D_LAMBDA	18,202	-0.001	0.014	-0.072	0.000	0.054
D_ETF_%HLD	18,202	0.004	0.006	-0.015	0.003	0.033
D_MFID_%HLD	18,202	0.000	0.002	-0.007	0.000	0.010
D_MFAM_%HLD	18,202	0.000	0.035	-0.110	0.000	0.108
D_SIZE	18,202	0.063	0.238	-0.762	0.057	0.796
D_BTM	18,202	-0.002	0.082	-0.268	-0.001	0.290
D_VOLUME	18,202	0.080	0.402	-0.999	0.068	1.309
D_STDRET	18,202	0.112	3.629	-8.972	0.078	9.383
D_ABS_ABNRET	18,202	-0.005	0.150	-0.518	-0.002	0.498
D_PRICE	18,202	0.022	0.223	-0.768	0.031	0.627
D_ANALYST	18,202	0.045	0.264	-0.693	0.000	1.099
D_INST_%	18,202	0.027	0.082	-0.203	0.016	0.362
D_INST_N	18,202	0.057	0.180	-0.330	0.033	1.079
D_PST	18,202	0.008	0.133	-0.533	0.005	0.598
D_LAMBDA_MKT	18,202	-0.004	0.029	-0.353	0.005	0.038
POST_SIZE	18,202	14.24	1.57	11.16	14.02	18.58
POST_BTM	18,202	0.652	0.262	0.132	0.663	1.334
POST_ANALYST	18,202	2.020	0.829	0.000	2.079	3.497
POST_INST_%	18,202	0.691	0.239	0.000	0.739	1.000
POST_INST_N	18,202	4.957	0.856	0.000	4.934	7.016
STK_DVF1	18,202	0.895	0.306	0.000	1.000	1.000
STK_DVF2	18,202	0.830	0.375	0.000	1.000	1.000

D_Xs refer to the change of variables calculated over one quarter pre- and post-periods surrounding quarters when an ETF is newly introduced. Let *t*=0 be the event quarter and *X(t)* a variable for a period that is *t* quarters before and after the event period. **D_X** is calculated as $\frac{1}{k}\sum_{t=1}^{k} X(t) - \frac{1}{k}\sum_{t=-k}^{-1} X(t)$, where *k*=1. The sample consists of firm-quarter observations with both (t-1) and (t+1) quarters available. **STK_DVF1** and **STK_DVF2** are calculated from the ETF-level variables, **ETF_DVF1** and **ETF_DVF2**. **ETF_DVF1** is 0 if an ETF is a sector ETF, and 1 otherwise. **ETF_DVF2** is 0(1) if an ETF is ranked in the lower (higher) 50% of the factor score, calculated with the diversification score, the number of holding stocks, and the average weights of holding stocks. The diversification score is 0 if an ETF is a sector ETF, 1 if an ETF duplicates an index based on size and growth, 2 if an ETF duplicates an index based either on size or growth, and 3 if an ETF duplicates a broad market index. For firms that are included in new ETFs, **ETF_DVF1** and **ETF_DVF2** of these new ETFs are weighted by the number of shares held by these ETFs. **STK_DVF1 (STK_DVF2)** is 0 if the value-weighted **ETF_DVF1 (ETF_DVF2)** for the firm-quarter is closer to 0, and 1 if it is closer to 1. Tabulated in panel B are **STK_DVF1, STK_DVF2**, and **D_X**s calculated for one quarter periods for the following variables. **LAMBDA** (in %) is the adverse selection component (λ) of bid-ask spreads estimated per month with the method by Madhavan et al. [1997], modified by Armstrong et al. [2009]. Observations with negative values or larger than 100(%) of **LAMBDA** are excluded from the sample. **ETF_%HLD** is the percentage of shares held by any of the U.S. equity ETFs calculated with Thomson holdings data (S12). **MFID_%HLD** is the percentage of shares held by U.S. equity index mutual funds calculated with Thomson holdings data (S12). **MFAM_%HLD** is the percentage of shares held by actively managed U.S. equity mutual funds. All non-ETF, non-index, U.S. equity mutual funds in Thomson holdings data (S12) are categorized as actively managed U.S. equity mutual funds. **ETF_%HLD, MFID_%HLD**, and **MFAM_%HLD** are calculated daily and averaged by quarter. **SIZE** is the log of one plus market value (CRSP *prc* * *shrout*). **BTM** is the book-to-market assets ratio calculated as (Compustat *ATQ*/(*ATQ* -Compustat *CEQQ* + Market Value))) on the quarter end date. **VOLUME** it the turnover calculated as the log of one plus the quarterly average of daily CRSP trading volume divided by one plus the number of shares outstanding. **STDRET** is the annualized time-series standard deviation of daily returns (CRSP *ret*), calculated each quarter. **ABS_ABNRET** is the absolute abnormal cumulative returns in each quarter. **PRICE** is the log of one plus the quarterly average of daily CRSP prices. **ANALYST** is the log of one plus the quarterly average number of analysts following counted from I/B/E/S. **INST_%** is the quarterly % of institutional ownership from Thomson Financial Spectrum. **INST_N** is the log of one plus the quarterly number of institutional owners from Thomson Financial Spectrum. **PST** is the % of small trades (defined as trades smaller than $5,000) calculated with the TAQ intraday trade dataset. **PST** is calculated daily and averaged by quarter. **LAMBDA_MKT** is the average **LAMBDA** of all stocks in the sample. All variables are winsorized at 1%.

Panel C: Firm-quarter observations of change variables over two, three, and four-quarter pre- & post-periods

Variable	N	Mean	Std Dev	Min	Median	Max
No. QTRs: 2						
D_LAMBDA	17,783	-0.002	0.016	-0.080	0.000	0.088
D_ETF_%HLD	17,783	0.006	0.007	-0.012	0.005	0.037
No. QTRs: 3						
D_LAMBDA	17,379	-0.003	0.020	-0.103	0.000	0.103
D_ETF_%HLD	17,379	0.007	0.008	-0.011	0.006	0.041
No. QTRs: 4						
D_LAMBDA	16,963	-0.005	0.023	-0.125	-0.001	0.106
D_ETF_%HLD	16,963	0.009	0.008	-0.010	0.007	0.044

D_Xs refer to the change of variables calculated over one to four quarters of pre- and post-periods surrounding quarters when an ETF is newly introduced. Let *t*=0 be the event quarter and *X(t)* a variable for a period that is *t* quarters before and after the event period. **D_X** is calculated as $\frac{1}{k}\sum_{t=1}^{k} X(t) - \frac{1}{k}\sum_{t=-k}^{-1} X(t)$,where *k*=2,..,4. Tabulated in panel C are **D_X**s calculated for two to four quarter periods for **LAMBDA** and **ETF_%HLD**. The sample used for the first two rows consists of firm-quarter observations with (t-2), (t-1), (t+1), and (t+2) quarters available. The sample used for the next two rows consists of firm-quarter observations with (t-3), (t-2), (t-1), (t+1), (t+2), and (t+3) quarters available. The sample used for the last two rows consists of firm-quarter observations with (t-4), (t-3), (t-2), (t-1), (t+1), (t+2), (t+3), and (t+4) quarters available. **LAMBDA** (in %) is the adverse selection component (λ) of bid-ask spreads estimated per month with the method by Madhavan et al. [1997], modified by Armstrong et al. [2009]. Observations with negative values or larger than 100(%) of **LAMBDA** are excluded from the sample. **ETF_%HLD** is the percentage of shares held by any of the U.S. equity ETFs calculated with Thomson holdings data (S12). The holding data are reported each month or quarter. All variables are winsorized at 1%.

Panel D: ETF-month observations

Variable	N	Mean	Std Dev	Min	Median	Max
ETF_LAMBDA	8,983	0.024	0.058	0.000	0.011	1.968
STK_LAMBDA	8,983	0.015	0.011	0.003	0.012	0.089
ETF_SIZE	8,983	2,012.69	6,303.28	1.12	270.66	96,723.11
ETF_VOLUME	8,983	-3.996	1.134	-8.690	-4.195	0.684
ETF_STDRET	8,981	0.244	0.189	0.000	0.185	1.472
ETF_ABS_ABNRET	8,983	0.023	0.027	0.000	0.014	0.344
ETF_PRICE	8,983	54.95	32.57	1.80	51.85	205.02
ETF_INST_%	8,983	0.380	0.276	0.000	0.328	1.000
ETF_INST_N	8,983	54.17	72.61	0.00	25.00	575.00
STK_IDIO	8,983	0.016	0.007	0.005	0.014	0.061
ETF_DVF1	8,983	0.520	0.500	0.000	1.000	1.000
ETF_DVF2	8,983	0.492	0.500	0.000	0.000	1.000
ETF_PST	8,983	0.221	0.258	0.000	0.000	0.968
ETF_LAMBDA_MKT	8,983	0.025	0.019	0.004	0.022	0.084

ETF_LAMBDA (in %) is the adverse selection component (λ) of ETF bid-ask spreads estimated per month with the method by Madhavan et al. [1997] modified by Armstrong et al. [2009]. Observations with negative values or larger than 100(%) of ETF_LAMBDA are excluded from the sample. **STK_LAMBDA** is the value-weighted average of holding stocks' lambdas of ETFs. **ETF_SIZE** the log of one plus ETF market value, calculated with the sum of stocks' daily market values (CRSP *prc* * *shrout*). Tabulated are unlogged values of **ETF_SIZE** in millions. **ETF_VOLUME** it the turnover of an ETF calculated as the log of one plus the quarterly average of daily CRSP trading volume divided by one plus the number of shares outstanding. **ETF_STDRET** is the annualized time-series standard deviation of daily returns, calculated each quarter. **ETF_ABS_ABNRET** is the absolute abnormal cumulative returns in each quarter. **ETF_PRICE** is the log of one plus the quarterly average of daily CRSP prices of an ETF. Tabulated are the unlogged values of **ETF_PRICE**. **ETF_INST_%** is the quarterly % of institutional ownership of an ETF from Thomson Financial Spectrum. **ETF_INST_N** is the log of one plus the quarterly number of institutional owners of an ETF from Thomson Financial Spectrum. Tabulated are the unlogged values of **ETF_INST_N. STK_IDIO** is the value-weighted average of holding stocks' idiosyncratic volatilities. Idiosyncratic volatility is calculated as the quarterly standard deviation of residuals of daily cross-sectional Fama-French 3-factor regressions. At least 15 daily observations are required to calculate the standard deviation for each quarter. **ETF_DVF1** is 0 if an ETF is a sector ETF, and 1 otherwise. **ETF_DVF2** is 0(1) if an ETF is ranked in the lower (higher) 50% of the factor score, calculated with the diversification score, the number of holding stocks, and the average weights of holding stocks. The diversification score is 0 if an ETF is a sector ETF, 1 if an ETF duplicates an index based on size and growth, 2 if an ETF duplicates an index based either on size or growth, and 3 if an ETF duplicates a broad market index. **ETF_PST** is the % of small trades, calculated with intraday trade data of ETFs. **ETF_LAMBDA_MKT** is the average of **ETF_LAMBDA,** calculated with all sample ETFs. All variables are winsorized at 1%.

Table 3: Pearson Correlations

Panel A: Firm-quarter observations

	ETF _%HLD	MFID _%HLD	MFAM _%HLD	SIZE	BTM	VOL UME	STD RET	ABS_ ABN RET	PRICE	ANA LYST	INST _%	INST _N	PST	LAM BDA _MKT
LAMBDA	-0.27	-0.22	-0.30	-0.55	0.19	-0.53	0.09	0.14	-0.44	-0.40	-0.38	-0.41	0.26	0.29
ETF_%HLD		0.42	0.47	0.43	-0.14	0.44	-0.02	-0.10	0.36	0.50	0.61	0.56	-0.12	-0.14
MFID_%HLD			0.45	0.60	-0.13	0.32	-0.03	-0.14	0.40	0.56	0.47	0.58	-0.43	-0.02
MFAM_%HLD				0.52	-0.19	0.43	0.00	-0.12	0.44	0.65	0.78	0.66	-0.31	-0.04
SIZE					-0.27	0.53	-0.04	-0.23	0.73	0.73	0.56	0.71	-0.63	-0.16
BTM						-0.26	0.00	-0.04	-0.12	-0.22	-0.11	-0.21	0.07	0.17
VOLUME							-0.09	0.11	0.28	0.55	0.54	0.50	-0.18	-0.14
STDRET								0.04	0.12	-0.04	-0.02	-0.02	-0.03	0.04
ABS_ABNRET									-0.33	-0.13	-0.14	-0.16	0.21	0.13
PRICE										0.50	0.50	0.55	-0.74	-0.18
ANALYST											0.70	0.80	-0.38	-0.08
INST_%												0.80	-0.30	-0.12
INST_N													-0.39	-0.11
PST														0.03

LAMBDA (in %) is the adverse selection component (λ) of bid-ask spreads estimated per month with the method by Madhavan et al. [1997], modified by Armstrong et al. [2009]. Observations with negative values or larger than 100(%) of **LAMBDA** are excluded from the sample. **ETF_%HLD** is the percentage of shares held by any of the U.S. equity ETFs, calculated with Thomson holdings data (S12). The holding data are reported each month or quarter. **MFID_%HLD** is the percentage of shares held by U.S. equity index mutual funds, calculated with Thomson holdings data (S12). The holding data are reported each month or quarter. **MFAM_%HLD** is the percentage of shares held by actively managed U.S. equity mutual funds. All non-ETF, non-index, U.S. equity mutual funds in Thomson holdings data (S12) are categorized as actively managed U.S. equity mutual funds. The holding data are reported each month or quarter. **ETF_%HLD, MFID_%HLD**, and **MFAM_%HLD** are calculated daily and averaged by quarter. **SIZE** is the log of one plus market value (CRSP *prc* * *shrout*). **BTM** is the book-to-market assets ratio calculated as (Compustat *ATQ*/(*ATQ* -Compustat *CEQQ* + Market Value))) on the quarter end date. **VOLUME** is the turnover calculated as the log of one plus the quarterly average of daily CRSP trading volume divided by one plus the number of shares outstanding. **STDRET** is the annualized time-series standard deviation of daily returns (CRSP *ret*), calculated each quarter. **ABS_ABNRET** is the absolute abnormal cumulative returns in each quarter. **PRICE** is the log of one plus the quarterly average of daily CRSP prices. **ANALYST** is the log of one plus the quarterly average number of analysts following counted from I/B/E/S. **INST_%** is the quarterly % of institutional ownership from Thomson Financial Spectrum. **INST_N** is the log of one plus the quarterly number of institutional owners from Thomson Financial Spectrum. **PST** is the % of small trades (defined as trades smaller than $5,000), calculated with the TAQ intraday trade dataset. **PST** is calculated daily and averaged by quarter. **LAMBDA_MKT** is the average **LAMBDA** of all stocks in the sample. All variables are winsorized at 1%.

Panel B: Change variables over one-quarter pre- and post-periods

	D_ ETF _%HLD	D_ MFID _%HLD	D_ MFAM _%HLD	D_ SIZE	D_ BTM	D_VOL UME	D_STD RET	D_ABS _ABN RET	D_ PRICE	D_ANA LYST	D_ INST_%	D_IN ST_N	D_ PST	D_LAM BDA_ MKT	STK_ DVF1	STK_ DVF2
D_LAMBDA	0.04	0.01	-0.07	-0.33	0.24	-0.25	0.00	0.06	-0.28	-0.12	-0.11	-0.29	0.10	0.21	-0.06	-0.05
D_ETF_%HLD		0.30	0.20	0.02	0.00	0.06	0.02	0.01	0.14	0.08	0.24	0.15	-0.12	0.08	-0.13	-0.05
D_MFID_%HLD			0.20	0.02	0.03	0.06	0.00	0.00	0.12	0.06	0.22	0.11	-0.08	-0.03	-0.01	0.00
D_MFAM_%HLD				0.11	-0.06	0.09	-0.02	-0.02	0.21	0.15	0.41	0.20	-0.09	-0.05	0.01	0.02
D_SIZE					-0.73	0.14	0.00	-0.11	0.85	0.16	0.20	0.54	-0.40	-0.30	0.02	0.00
D_BTM						-0.13	-0.01	0.06	-0.68	-0.06	-0.11	-0.35	0.32	0.24	-0.04	-0.01
D_VOLUME							0.07	0.19	0.17	0.11	0.23	0.30	-0.14	0.04	-0.02	-0.02
D_STDRET								0.06	0.02	0.00	0.00	0.00	-0.01	0.03	0.00	0.01
D_ABS_ABNRET									-0.08	0.00	-0.03	-0.03	-0.01	0.10	-0.04	-0.03
D_PRICE										0.10	0.26	0.46	-0.54	-0.27	0.02	0.00
D_ANALYST											0.19	0.29	-0.02	-0.07	0.00	0.02
D_INST_%												0.45	-0.11	-0.03	0.03	0.06
D_INST_N													-0.21	-0.16	0.01	0.03
D_PST														0.13	-0.01	0.00
D_LAMBDA_MKT															-0.19	-0.14
STK_DVF1																0.69

D_Xs refer to the change of variables calculated over one to four quarters of pre- and post-periods surrounding quarters when an ETF is newly introduced. Let t=0 be the event quarter and $X(t)$ denote a variable for a period that is t quarters before and after the event period. **D_X** is calculated as $\frac{1}{k}\sum_{t=1}^{k}X(t)-\frac{1}{k}\sum_{t=-k}^{-1}X(t)$, where k=1,…,4.

STK_DVF1 and **STK_DVF2** are calculated from the ETF-level variables, **ETF_DVF1** and **ETF_DVF2**. **ETF_DVF1** is 0 if an ETF is a sector ETF, and 1 otherwise. **ETF_DVF2** is 0(1) if an ETF is ranked in the lower (higher) 50% of the factor score, calculated with the diversification score, the number of holding stocks, and the average weights of holding stocks. The diversification score is 0 if an ETF is a sector ETF, 1 if an ETF duplicates an index based on size and growth, 2 if an ETF duplicates an index based either on size or growth, and 3 if an ETF duplicates a broad market index. For firms that are included in new ETFs, **ETF_DVF1** and **ETF_DVF2** of these new ETFs are weighted by the number of shares held by these ETFs. **STK_DVF1 (STK_DVF2)** is 0 if the value-weighted **ETF_DVF1 (ETF_DVF2)** for the firm-quarter is closer to 0, and 1 if it is closer to 1.All variables are winsorized at 1%.

Panel C: ETF-month observations

	STK_ LAM BDA	ETF_ SIZE	ETF_ VOL UME	ETF_ STD RET	ETF_ ABS_ ABNRET	ETF_ PRICE	ETF_ INST _%	ETF_ INST _N	STK_ IDIO	ETF_ DVF1	ETF_ DVF2	ETF_ PST	ETF_L AMBD A_MKT
ETF_LAMBDA	0.21	-0.32	0.01	0.28	0.14	-0.24	-0.03	-0.24	0.29	-0.03	-0.02	0.17	0.30
STK_LAMBDA		-0.17	0.15	0.61	0.25	-0.20	0.05	-0.02	0.77	0.12	0.16	0.21	0.52
ETF_SIZE			0.10	-0.17	-0.14	0.54	0.26	0.79	-0.23	0.21	0.19	-0.37	-0.20
ETF_VOLUME				0.28	0.23	0.08	0.52	0.22	0.23	-0.27	-0.23	0.03	0.17
ETF_STDRET					0.40	-0.30	0.15	0.02	0.84	-0.14	-0.09	0.34	0.75
ETF_ABS_ABNRET						-0.18	0.12	-0.04	0.40	-0.35	-0.30	0.21	0.27
ETF_PRICE							0.02	0.36	-0.31	0.23	0.30	-0.82	-0.24
ETF_INST_%								0.45	0.10	-0.28	-0.20	0.08	0.05
ETF_INST_N									-0.03	0.04	0.08	-0.18	-0.02
STK_IDIO										-0.02	0.04	0.31	0.70
ETF_DVF1											0.79	-0.27	-0.02
ETF_DVF2												-0.33	0.02
ETF_PST													0.25

ETF_LAMBDA (in %) is the adverse selection component (λ) of ETF bid-ask spreads estimated per month with the method by Madhavan et al. [1997] modified by Armstrong et al. [2009]. Observations with negative values or larger than 100(%) of ETF_LAMBDA are excluded from the sample. **STK_LAMBDA** is the value-weighted average of holding stocks' lambdas of ETFs. **ETF_SIZE** the log of one plus ETF market value, calculated with the sum of stocks' daily market values (CRSP *prc* * *shrout*). **ETF_VOLUME** it the turnover of an ETF calculated as the log of one plus the quarterly average of daily CRSP trading volume divided by one plus the number of shares outstanding. **ETF_STDRET** is the annualized time-series standard deviation of daily returns, calculated each quarter. **ETF_ABS_ABNRET** is the absolute abnormal cumulative returns in each quarter. **ETF_PRICE** is the log of one plus the quarterly average of daily CRSP prices of an ETF. **ETF_INST_%** is the quarterly % of institutional ownership of an ETF from Thomson Financial Spectrum. **ETF_INST_N** is the log of one plus the quarterly number of institutional owners of an ETF from Thomson Financial Spectrum. **STK_IDIO** is the value-weighted average of holding stocks' idiosyncratic volatilities. Idiosyncratic volatility is calculated as the quarterly standard deviation of residuals of daily cross-sectional Fama-French 3-factor regressions. At least 15 daily observations are required to calculate the standard deviation for each quarter. **ETF_DVF1** is 0 if an ETF is a sector ETF, and 1 otherwise. **ETF_DVF2** is 0(1) if an ETF is ranked in the lower (higher) 50% of the factor score, calculated with the diversification score, the number of holding stocks, and the average weights of holding stocks. The diversification score is 0 if an ETF is a sector ETF, 1 if an ETF duplicates an index based on size and growth, 2 if an ETF duplicates an index based either on size or growth, and 3 if an ETF duplicates a broad market index. **ETF_PST** is the % of small trades, calculated with intraday trade data of ETFs. **ETF_LAMBDA_MKT** is the average of **ETF_LAMBDA,** calculated with all sample ETFs. All variables are winsorized at 1%.

Table 4: Regressions of Adverse Selection

Panel A: Regression of LAMBDA on % of shares held by ETFs

	Pred. Sign	Dep. Var: LAMBDA			
		1	2	3	4
Intercept		0.268 ***	0.238 ***	0.133 **	0.603 ***
		(4.18)	(4.15)	(2.09)	(5.87)
ETF_%HLD	**+**		**1.362 ****	**1.217 ****	**1.739 *****
			(2.66)	**(2.16)**	**(2.97)**
SIZE	**-**	-0.043 ***	-0.042 ***	-0.042 ***	-0.053 ***
		(-7.8)	(-7.84)	(-7.85)	(-8.56)
BTM	**+**	-0.007	-0.005	-0.012 **	-0.020 ***
		(-1.15)	(-0.88)	(-2.46)	(-3.48)
VOLUME	-	-0.078 ***	-0.080 ***	-0.080 ***	-0.078 ***
		(-12.49)	(-12.53)	(-12.33)	(-12.47)
STDRET	+	0.005 ***	0.005 ***	0.005 ***	0.006 ***
		(3.47)	(3.66)	(3.91)	(4.66)
ABS_ABNRET	+	0.114 ***	0.116 ***	0.117 ***	0.096 ***
		(8.45)	(8.37)	(7.5)	(6.72)
PRICE	+/-	-0.054 ***	-0.056 ***	-0.053 ***	-0.097 ***
		(-5.12)	(-5.19)	(-4.96)	(-6.28)
ANALYST	-	0.036 ***	0.035 ***	0.034 ***	0.036 ***
		(8.17)	(8.28)	(7.98)	(8.91)
INST_%	+	0.059 ***	0.033 ***	0.037 ***	0.043 ***
		(4.43)	(3.13)	(3.29)	(3.95)
INST_N	-	-0.012 ***	-0.014 ***	-0.014 ***	-0.015 ***
		(-5.52)	(-5.62)	(-6.64)	(-7.45)
LAMBDA_MKT	+			0.608 ***	0.591 ***
				(9.77)	(10.04)
PST	-				-0.263 ***
					(-7.53)
R-sqr		*0.437*	*0.440*	*0.445*	*0.461*

*, **, *** Significant at the 0.10, 0.05, and 0.01 level, respectively. This table provides regressions of the adverse selection components of all Compustat firms on the percentages of shares held by ETFs. I estimate pooled regressions of the firm sample and report coefficients and t-statistics calculated with two-way cluster robust errors (Gow et al. [2009]). **LAMBDA** (in %) is the adverse selection component (λ) of bid-ask spreads estimated per month with the method by Madhavan et al. [1997], modified by Armstrong et al. [2009]. Observations with negative values or larger than 100(%) of **LAMBDA** are excluded from the sample. **ETF_%HLD** is the percentage of shares held by any of the U.S. equity ETFs, calculated with Thomson holdings data (S12). The holding data are reported each month or quarter. **SIZE** is the log of one plus market value (CRSP *prc* * *shrout*). **BTM** is the book-to-market assets ratio calculated as (Compustat *ATQ*/(*ATQ* -Compustat *CEQQ* + Market Value))) on the quarter end date. **VOLUME** is the turnover calculated as the log of one plus the quarterly average of daily CRSP trading volume divided by one plus the number of shares outstanding. **STDRET** is the annualized time-series standard deviation of daily returns (CRSP *ret*), calculated each quarter. **ABS_ABNRET** is the absolute abnormal cumulative returns in each quarter. **PRICE** is the log of one plus the quarterly average of daily CRSP prices. **ANALYST** is the log of one plus the quarterly average number of analysts following counted from I/B/E/S. **INST_%** is the quarterly % of institutional ownership from Thomson Financial Spectrum. **INST_N** is the log of one plus the quarterly number of institutional owners from Thomson Financial Spectrum. **LAMBDA_MKT** is the average **LAMBDA** of all stocks in the sample. **PST** is the % of small trades (defined as trades smaller than $5,000), calculated with the TAQ intraday trade dataset. **PST** is calculated daily and averaged by quarter. All variables are winsorized at 1%.

Panel B: Regression of LAMBDA on % of shares held by mutual funds

	Pred. Sign	Dep. Var: LAMBDA			
		1	2	3	4
Intercept		0.676 ***	0.674 ***	0.676 ***	0.635 ***
		(6.47)	(6.52)	(6.55)	(6)
ETF_%HLD	+		**1.295 ****	**1.291 ****	
			(2.27)	**(2.26)**	
MFID_%HLD	+	**6.111 *****	**5.398 *****	**5.356 *****	
		(7.71)	**(7.69)**	**(7.7)**	
ID_%HLD	+				**2.005 *****
					(3.78)
MFAM_%HLD	+			0.062 *	0.075 **
				(1.91)	(2.14)
SIZE	-	-0.062 ***	-0.061 ***	-0.061 ***	-0.056 ***
		(-8.84)	(-8.89)	(-8.9)	(-8.4)
BTM	+	-0.024 ***	-0.022 ***	-0.021 ***	-0.018 ***
		(-4.42)	(-4.08)	(-3.74)	(-3.14)
VOLUME	-	-0.074 ***	-0.076 ***	-0.076 ***	-0.077 ***
		(-12.44)	(-12.59)	(-12.59)	(-12.44)
STDRET	+	0.005 ***	0.005 ***	0.005 ***	0.006 ***
		(4.2)	(4.53)	(4.55)	(4.56)
ABS_ABNRET	+	0.099 ***	0.099 ***	0.099 ***	0.097 ***
		(6.92)	(6.87)	(6.87)	(6.77)
PRICE	+/-	-0.080 ***	-0.086 ***	-0.086 ***	-0.095 ***
		(-6.01)	(-5.91)	(-5.93)	(-6.4)
ANALYST	-	0.034 ***	0.034 ***	0.033 ***	0.034 ***
		(8.91)	(8.98)	(8.44)	(8.59)
INST_%	+	0.066 ***	0.043 ***	0.030 ***	0.020 *
		(5.61)	(4.06)	(2.94)	(1.86)
INST_N	-	-0.019 ***	-0.020 ***	-0.019 ***	-0.017 ***
		(-9.33)	(-9.56)	(-9.08)	(-7.32)
LAMBDA_MKT	+	0.620 ***	0.590 ***	0.589 ***	0.579 ***
		(10.31)	(9.82)	(9.84)	(9.83)
PST	-	-0.209 ***	-0.229 ***	-0.229 ***	-0.256 ***
		(-7.25)	(-7.3)	(-7.32)	(-7.54)
R-sqr		*0.464*	*0.467*	*0.467*	*0.464*

*, **, *** Significant at the 0.10, 0.05, and 0.01 level, respectively. This table provides regressions of the adverse selection components of all Compustat firms on the percentages of shares held by ETFs and the percentages of shares held by mutual funds. I estimate pooled regressions of the firm sample and report coefficients and t-statistics calculated with two-way cluster robust errors (Gow et al. [2009]). **LAMBDA** (in %) is the adverse selection component (λ) of bid-ask spreads estimated per month with the method by Madhavan et al. [1997], modified by Armstrong et al. [2009]. Observations with negative values or larger than 100(%) of **LAMBDA** are excluded from the sample. **ETF_%HLD** is the percentage of shares held by any of the U.S. equity ETFs, calculated with Thomson holdings data (S12). The holding data are reported each month or quarter. **MFID_%HLD** is the percentage of shares held by U.S. equity index mutual funds, calculated with Thomson holdings data (S12). The holding data are reported each month or quarter. **MFAM_%HLD** is the percentage of shares held by actively managed U.S. equity mutual funds. All non-ETF, non-index, U.S. equity mutual funds in

Thomson holdings data (S12) are categorized as actively managed U.S. equity mutual funds. The holding data are reported each month or quarter. **ETF_%HLD, MFID_%HLD**, and **MFAM_%HLD** are calculated daily and averaged by quarter. **ID_%HLD** is the sum of **ETF_%HLD** and **MFID_%HLD. SIZE** is the log of one plus market value (CRSP *prc* * *shrout*). **BTM** is the book-to-market assets ratio calculated as (Compustat *ATQ*/(*ATQ* - Compustat *CEQQ* + Market Value))) on the quarter end date. **VOLUME** is the turnover calculated as the log of one plus the quarterly average of daily CRSP trading volume divided by one plus the number of shares outstanding. **STDRET** is the annualized time-series standard deviation of daily returns (CRSP *ret*), calculated each quarter. **ABS_ABNRET** is the absolute abnormal cumulative returns in each quarter. **PRICE** is the log of one plus the quarterly average of daily CRSP prices. **ANALYST** is the log of one plus the quarterly average number of analysts following counted from I/B/E/S. **INST_%** is the quarterly % of institutional ownership from Thomson Financial Spectrum. **INST_N** is the log of one plus the quarterly number of institutional owners from Thomson Financial Spectrum. **LAMBDA_MKT** is the average **LAMBDA** of all stocks in the sample. **PST** is the % of small trades (defined as trades smaller than $5,000), calculated with the TAQ intraday trade dataset. **PST** is calculated daily and averaged by quarter. All variables are winsorized at 1%.

Table 5: Regressions of the Change in Adverse Selection

Panel A: Regressions of D_LAMBDA on D_ETF_%HLD and change variables

		Dep. Var: D_LAMBDA							
	Pred. Sign	1 QTR				2 QTR			
		1	2	3	4	5	6	7	8
Intercept		0.001	0.001	0.000	0.000	0.002*	0.002*	0.002*	0.002*
		(0.63)	(0.65)	(0.61)	(0.58)	(1.83)	(1.94)	(1.9)	(1.79)
D_ETF_%HLD	+	**0.131***	**0.118***	**0.121***		**0.161****	**0.140****	**0.140****	
		(1.82)	**(1.7)**	**(1.71)**		**(2.21)**	**(2.03)**	**(2.02)**	
D_MFID_%HLD	+		**0.170*****	**0.181*****			**0.313*****	**0.314*****	
			(3.97)	**(4.07)**			**(3.97)**	**(3.95)**	
D_ID_%HLD	+				**0.129****				**0.164****
					(2.03)				**(2.47)**
D_MFAM_%HLD	+/-			-0.007*	-0.007*			-0.001	-0.001
				(-1.94)	(-1.86)			(-0.32)	(-0.26)
D_SIZE	-	-0.010***	-0.010***	-0.010***	-0.010***	-0.014***	-0.014***	-0.014***	-0.014***
		(-3.39)	(-3.37)	(-3.46)	(-3.43)	(-5.64)	(-5.65)	(-5.7)	(-5.64)
D_BTM	+	-0.003	-0.003	-0.003	-0.003	-0.001	-0.002	-0.002	-0.001
		(-0.79)	(-0.89)	(-0.86)	(-0.85)	(-0.13)	(-0.26)	(-0.26)	(-0.21)
D_VOLUME	-	-0.007***	-0.007***	-0.007***	-0.007***	-0.011***	-0.011***	-0.011***	-0.011***
		(-7.45)	(-7.45)	(-7.46)	(-7.45)	(-5.64)	(-5.63)	(-5.63)	(-5.65)
D_STDRET	+	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		(0.1)	(0.12)	(0.09)	(0.08)	(1.38)	(1.37)	(1.36)	(1.37)
D_ABS_ABNRET	+	0.005***	0.005***	0.005***	0.005***	0.012***	0.012***	0.012***	0.012***
		(5.31)	(5.2)	(5.17)	(5.21)	(5.63)	(5.56)	(5.57)	(5.63)
D_PRICE	+/-	-0.005***	-0.005***	-0.005***	-0.005***	-0.003*	-0.003*	-0.003**	-0.003*
		(-2.84)	(-3.11)	(-3.2)	(-3.09)	(-1.65)	(-1.95)	(-2.01)	(-1.88)
D_ANALYST	-	-0.002***	-0.002***	-0.002***	-0.002***	-0.002**	-0.002**	-0.002**	-0.002**
		(-3)	(-3.01)	(-2.98)	(-2.98)	(-2.4)	(-2.38)	(-2.32)	(-2.34)
D_INST_%	+	0.004	0.004	0.005	0.005	0.000	-0.001	-0.001	-0.001
		(0.7)	(0.61)	(0.74)	(0.76)	(-0.05)	(-0.23)	(-0.15)	(-0.13)
D_INST_N	-	-0.008**	-0.008**	-0.008**	-0.008**	-0.005**	-0.005**	-0.005**	-0.005**

		(-2.28)	(-2.29)	(-2.3)	(-2.3)	(-2.1)	(-2.21)	(-2.18)	(-2.14)
D_LAMBDA_MKT	+	0.062***	0.063***	0.063***	0.062***	0.068**	0.069**	0.069**	0.068**
		(2.75)	(2.82)	(2.81)	(2.76)	(2.37)	(2.47)	(2.47)	(2.41)
D_PST	-	-0.007***	-0.007***	-0.007***	-0.007***	-0.010***	-0.010***	-0.010***	-0.010***
		(-3.56)	(-3.6)	(-3.6)	(-3.59)	(-2.93)	(-2.98)	(-2.99)	(-2.95)
R-sqr		*0.184*	*0.184*	*0.184*	*0.184*	*0.276*	*0.278*	*0.278*	*0.277*

Panel A (continued): Regressions of D_LAMBDA on D_ETF_%HLD and change variables

		Dep. Var: D_LAMBDA							
	Pred. Sign	3 QTR				4 QTR			
		9	10	11	12	13	14	15	16
Intercept		0.003***	0.003***	0.003***	0.003***	0.004***	0.004***	0.004***	0.004***
		(2.87)	(3.06)	(3.01)	(2.86)	(4.13)	(4.46)	(4.29)	(4.1)
D_ETF_%HLD	+	0.216**	0.187*	0.187*		0.278**	0.246**	0.246**	
		(2.08)	(1.91)	(1.9)		(2.2)	(2.04)	(2.03)	
D_MFID_%HLD	+		**0.442*** **	**0.442*** **			**0.491*** **	**0.491*** **	
			(4.15)	**(4.15)**			**(4.05)**	**(4.07)**	
D_ID_%HLD	+				**0.222** **				**0.280** **
					(2.36)				**(2.47)**
D_MFAM_%HLD	+/-			0.003	0.003			0.001	0.001
				(0.48)	(0.48)			(0.11)	(0.07)
D_SIZE	-	-0.016***	-0.016***	-0.016***	-0.016***	-0.016***	-0.016***	-0.016***	-0.016***
		(-6.04)	(-6.07)	(-6.07)	(-6.02)	(-6.43)	(-6.47)	(-6.47)	(-6.42)
D_BTM	+	0.001	-0.001	-0.001	0.000	0.001	0.000	0.000	0.000
		(0.07)	(-0.09)	(-0.09)	(-0.02)	(0.11)	(-0.04)	(-0.04)	(0.01)
D_VOLUME	-	-0.015***	-0.015***	-0.015***	-0.015***	-0.019***	-0.019***	-0.019***	-0.019***
		(-4.76)	(-4.75)	(-4.74)	(-4.79)	(-4.52)	(-4.51)	(-4.51)	(-4.56)
D_STDRET	+	0.000**	0.000**	0.000**	0.000**	0.000**	0.000**	0.000**	0.000**
		(2.03)	(2.03)	(2.04)	(2.04)	(2.06)	(2.09)	(2.09)	(2.07)
D_ABS_ABNRET	+	0.027***	0.026***	0.026***	0.026***	0.042***	0.042***	0.042***	0.042***
		(4.35)	(4.32)	(4.32)	(4.34)	(5.43)	(5.44)	(5.43)	(5.45)
D_PRICE	+/-	-0.003**	-0.004***	-0.004***	-0.004***	-0.004**	-0.004***	-0.004***	-0.004***

		(-2.32)	(-2.79)	(-2.85)	(-2.63)	(-2.38)	(-2.77)	(-2.8)	(-2.61)
D_ANALYST	-	-0.003***	-0.003***	-0.003***	-0.003***	-0.002**	-0.002**	-0.002**	-0.002**
		(-2.74)	(-2.7)	(-2.67)	(-2.69)	(-2.55)	(-2.47)	(-2.39)	(-2.41)
D_INST_%	+	-0.002	-0.003	-0.003	-0.003	0.002	0.002	0.001	0.001
		(-0.44)	(-0.6)	(-0.64)	(-0.65)	(0.35)	(0.31)	(0.21)	(0.18)
D_INST_N	-	-0.004**	-0.004**	-0.004**	-0.004**	-0.004**	-0.005**	-0.005**	-0.005**
		(-1.96)	(-2.19)	(-2.13)	(-2)	(-1.99)	(-2.25)	(-2.19)	(-2.03)
D_LAMBDA_MKT	+	0.091***	0.092***	0.092***	0.091***	0.102***	0.103***	0.103***	0.102***
		(2.77)	(2.87)	(2.87)	(2.81)	(2.77)	(2.84)	(2.84)	(2.81)
D_PST	-	-0.015***	-0.016***	-0.016***	-0.015***	-0.021**	-0.021***	-0.021***	-0.021**
		(-2.99)	(-3.06)	(-3.06)	(-3)	(-2.56)	(-2.61)	(-2.62)	(-2.57)
R-sqr		*0.363*	*0.366*	*0.366*	*0.365*	*0.438*	*0.441*	*0.441*	*0.441*

*, **, *** Significant at the 0.10, 0.05, and 0.01 level, respectively. This table provides regressions of the change in the adverse selection components of Compustat firms on the change in the percentages of those firms' shares that are held by ETFs. I estimate pooled regressions of the firm sample and report coefficients and t-statistics calculated with two-way cluster robust errors (Gow et al. [2009]). **D_X**s refer to the change of variables calculated over one to four quarters of pre- and post-periods surrounding quarters when an ETF is newly introduced. Let t=0 be the event quarter and *X(t)* denote a variable for a period that is t quarters before and after the event period. **D_X** is calculated as $\frac{1}{k}\sum_{t=1}^{k} X(t) - \frac{1}{k}\sum_{t=-k}^{-1} X(t)$,where k=1,…,4. All variables are winsorized at 1%. All variables are winsorized at 1%.

Panel B: Regressions of D_LAMBDA on D_ETF_%HLD, change variables, and post-period firm characteristics

	Pred. Sign	1 QTR		2 QTR		3 QTR		4 QTR	
		1	2	3	4	5	6	7	8
Intercept		-0.013*	-0.013*	-0.019**	-0.019**	-0.026**	-0.026**	-0.037***	-0.037***
		(-1.74)	(-1.7)	(-2.09)	(-2.03)	(-2.26)	(-2.19)	(-2.71)	(-2.64)
D_ETF_%HLD	+	**0.133****	**0.120***	**0.134*****	**0.121****	**0.158****	**0.146****	**0.204****	**0.199****
		(2.18)	**(1.89)**	**(2.7)**	**(2.3)**	**(2.41)**	**(2.05)**	**(2.41)**	**(2.14)**
D_MFID_%HLD	+		**0.139****		**0.149*****		**0.144**		**0.059**
			(2.31)		**(2.81)**		**(1.64)**		**(0.51)**
POST_SIZE	-	0.000	0.000	0.000	0.000	0.001	0.001	0.001*	0.001*
		(-0.28)	(-0.31)	(0.47)	(0.41)	(1.11)	(1.05)	(1.86)	(1.82)
POST_BTM	+	0.002	0.002	0.003**	0.003**	0.005***	0.005***	0.006***	0.006***
		(1.5)	(1.5)	(2.26)	(2.26)	(2.66)	(2.66)	(3.21)	(3.21)
D_VOLUME	-	-0.008***	-0.008***	-0.012***	-0.012***	-0.016***	-0.016***	-0.020***	-0.020***
		(-6.05)	(-6.09)	(-5.75)	(-5.76)	(-5.09)	(-5.09)	(-4.98)	(-4.97)
D_STDRET	+	0.000	0.000	0.000	0.000	0.000*	0.000*	0.000*	0.000*
		(0.34)	(0.36)	(1.37)	(1.37)	(1.84)	(1.85)	(1.74)	(1.75)
D_ABS_ABNRET	+	0.006***	0.006***	0.013***	0.013***	0.028***	0.028***	0.043***	0.043***
		(5.76)	(5.7)	(7.87)	(7.86)	(5.32)	(5.33)	(6.97)	(6.99)
D_PRICE	+/-	-0.016***	-0.016***	-0.018***	-0.018***	-0.021***	-0.021***	-0.023***	-0.023***
		(-5.84)	(-5.88)	(-6.58)	(-6.63)	(-7.15)	(-7.22)	(-7.24)	(-7.31)
POST_ANALYST	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		(0.42)	(0.42)	(0)	(0.01)	(0.33)	(0.32)	(0.04)	(0.04)
POST_INST_%	+	0.005**	0.005**	0.005	0.005	0.006*	0.006*	0.007*	0.007*
		(2.07)	(2.04)	(1.62)	(1.6)	(1.67)	(1.65)	(1.86)	(1.85)
POST_INST_N	-	0.002**	0.002**	0.002**	0.002***	0.002**	0.002**	0.003***	0.003***
		(2.35)	(2.36)	(2.55)	(2.58)	(1.98)	(2)	(2.88)	(2.89)
D_LAMBDA_MKT	+	0.058***	0.059***	0.069***	0.069***	0.092***	0.092***	0.100***	0.100***
		(3.35)	(3.42)	(3.23)	(3.28)	(3.61)	(3.64)	(3.63)	(3.65)

D_PST	-	-0.008***	-0.008***	-0.013***	-0.013***	-0.018***	-0.018***	-0.024***	-0.024***
		(-3.89)	(-3.91)	(-3.48)	(-3.5)	(-3.73)	(-3.77)	(-3.29)	(-3.33)
R-sqr		*0.189*	*0.190*	*0.274*	*0.275*	*0.361*	*0.361*	*0.447*	*0.447*

*, **, *** Significant at the 0.10, 0.05, and 0.01 level, respectively. This table provides regressions of the change in the adverse selection components of Compustat firms on the change in the percentages of those firms' shares that are held by ETFs. I estimate pooled regressions of the firm sample and report coefficients and t-statistics calculated with two-way cluster robust errors (Gow et al. [2009]). **D_X**s refer to the change of variables calculated over one to four quarters of pre- and post-periods surrounding quarters when an ETF is newly introduced. Let t=0 be the event quarter and *X(t)* denote a variable for a period that is t quarters before and after the event period. **D_X** is calculated as $\frac{1}{k}\sum_{t=1}^{k} X(t) - \frac{1}{k}\sum_{t=-k}^{-1} X(t)$,where k=1,…,4. **POST_X**s refer to the post period average of variables, calculated as $\frac{1}{k}\sum_{t=1}^{k} X(t)$,where k=1,…,4. All variables are winsorized at 1%.

Table 6: Regressions of the Change in Adverse Selection on Diversification

	Pred. Sign	Dep. Var: D_LAMBDA							
		1 QTR		2 QTR		3 QTR		4QTR	
		STK_DVF=							
		STK_DVF1	STK_DVF2	STK_DVF1	STK_DVF2	STK_DVF1	STK_DVF2	STK_DVF1	STK_DVF2
		1	2	3	4	5	6	7	8
Intercept		0.000	0.001 **	0.002 *	0.002 ***	0.003 ***	0.003 ***	0.004 ***	0.004 ***
		(0.48)	(2.11)	(1.74)	(2.83)	(3.1)	(3.6)	(2.91)	(3.9)
STK_DVF		0.000	-0.001	0.000	0.000	0.000	0.000	0.000	0.000
		(-0.11)	(-1.01)	(-0.25)	(-0.39)	(0.04)	(0.28)	(0.1)	(-0.14)
D_ETF_%HLD		-0.001	0.026	0.042	0.070 *	0.049	0.087 *	0.038	0.062 **
		(-0.03)	(0.45)	(1.23)	(1.86)	(0.98)	(1.91)	(1.58)	(2.33)
D_ETF_%HLD* STK_DVF	+	**0.175 ***	**0.155 ***	**0.148**	**0.131**	**0.203**	**0.175**	**0.288 ***	**0.284 *****
		(1.77)	**(1.7)**	**(1.46)**	**(1.37)**	**(1.45)**	**(1.33)**	**(1.92)**	**(2)**
D_SIZE		-0.011 ***	-0.009 ***	-0.014 ***	-0.010 ***	-0.013 ***	-0.010 ***	-0.015 ***	-0.012 ***
		(-4.55)	(-4.18)	(-7.53)	(-5.07)	(-7.32)	(-5.32)	(-7.41)	(-6.32)
D_BTM		-0.014 ***	-0.007 *	-0.013 ***	-0.008 *	-0.010 **	-0.006	-0.010 ***	-0.007 *
		(-4.72)	(-1.7)	(-5.57)	(-1.8)	(-2.27)	(-1.27)	(-3.25)	(-1.75)
D_VOLUME		-0.002 **	-0.002 **	-0.004 ***	-0.005 ***	-0.006 ***	-0.006 ***	-0.007 ***	-0.007 ***
		(-2.32)	(-2.31)	(-2.87)	(-3.35)	(-2.83)	(-3.32)	(-3.14)	(-4.16)
D_STDRET		0.000	0.000	0.000	0.000	0.000	0.000	0.000 *	0.000
		(1.46)	(0.94)	(0.3)	(-0.09)	(0.57)	(0.21)	(1.8)	(1.34)
D_ABS_ABNRET		0.004	0.002	0.009 *	0.008 **	0.019 ***	0.016 ***	0.025 ***	0.023 ***
		(1.17)	(0.89)	(1.92)	(2.13)	(2.6)	(2.87)	(2.96)	(3.93)
D_PRICE		-0.002	-0.002	0.000	-0.001	-0.001	-0.002	-0.002	-0.003
		(-1.22)	(-1.63)	(0.07)	(-1.25)	(-0.72)	(-1.25)	(-0.93)	(-1.51)
D_ANALYST		-0.001	-0.001 *	-0.002 *	-0.002 **	-0.002	-0.002	-0.002	-0.002
		(-1.54)	(-1.83)	(-1.91)	(-2.27)	(-1.15)	(-1.54)	(-1.26)	(-1.49)
D_INST_%		-0.002	-0.003	-0.010 ***	-0.007 **	-0.010 **	-0.010 ***	-0.013 ***	-0.011 ***
		(-0.46)	(-0.72)	(-3.55)	(-2.14)	(-2.48)	(-3.63)	(-2.62)	(-3.62)
D_INST_N		0.000	0.003	0.003 ***	0.002 *	0.002 ***	0.002 *	0.003 ***	0.002 **
		(-0.27)	(1.17)	(2.88)	(1.93)	(3.05)	(1.84)	(3.29)	(2.19)

D_LAMBDA_MKT	0.143 ***	0.055 ***	0.110 **	0.066 ***	0.100 ***	0.076 ***	0.107 ***	0.079 ***
	(3.52)	(3.06)	(2.1)	(5.04)	(3.63)	(8.66)	(5.59)	(7.13)
D_PST	-0.002 **	-0.002	-0.003 *	-0.002	-0.005 **	-0.004 *	-0.007 *	-0.006 *
	(-2.24)	(-1.57)	(-1.72)	(-1.42)	(-2.04)	(-1.65)	(-1.84)	(-1.89)
D_SIZE * STK_DVF	0.001	-0.001	-0.001	-0.005	-0.003	-0.007 **	-0.002	-0.006
	(0.33)	(-0.15)	(-0.14)	(-1.62)	(-0.9)	(-2.11)	(-0.54)	(-1.59)
D_BTM * STK_DVF	0.012 ***	0.003	0.012 **	0.006	0.011	0.005	0.011	0.007
	(2.93)	(0.72)	(2.22)	(1.18)	(1.17)	(0.59)	(1.26)	(1.06)
D_VOLUME * STK_DVF	-0.006 ***	-0.006 ***	-0.007 ***	-0.007 ***	-0.009 ***	-0.009 ***	-0.013 ***	-0.013 ***
	(-4.18)	(-4.51)	(-3.17)	(-3.34)	(-2.73)	(-3.07)	(-2.74)	(-3.11)
D_STDRET * STK_DVF	0.000 ***	0.000 *	0.000	0.000	0.000	0.000 *	0.000	0.000
	(-3.22)	(-1.76)	(0.83)	(1.42)	(0.73)	(1.79)	(0.18)	(1.04)
D_ABS_ABNRET * STK_DVF	0.002	0.003	0.002	0.004	0.008	0.011	0.018	0.021 **
	(0.48)	(1.25)	(0.42)	(0.82)	(0.83)	(1.39)	(1.57)	(2.22)
D_PRICE * STK_DVF	-0.003	-0.004	-0.004	-0.002	-0.002	-0.002	-0.002	-0.001
	(-1.04)	(-1.42)	(-1.62)	(-1.01)	(-1)	(-0.85)	(-0.77)	(-0.29)
D_ANALYST * STK_DVF	-0.001	-0.001	0.000	0.000	-0.001	-0.001	0.000	0.000
	(-0.46)	(-0.56)	(-0.07)	(0)	(-0.88)	(-0.41)	(-0.29)	(0.06)
D_INST_% * STK_DVF	0.007	0.009	0.010 *	0.008	0.009	0.009 **	0.015 **	0.015 ***
	(0.84)	(1.12)	(1.83)	(1.5)	(1.6)	(2.21)	(2.57)	(3.28)
D_INST_N * STK_DVF	-0.009 **	-0.012 **	-0.008 ***	-0.008 ***	-0.007 ***	-0.006 ***	-0.008 ***	-0.008 ***
	(-1.97)	(-2.52)	(-3.43)	(-3.09)	(-2.82)	(-2.61)	(-2.92)	(-2.74)
D_LAMBDA_MKT * STK_DVF	-0.087	0.004	-0.047	-0.002	-0.011	0.015	-0.005	0.024
	(-1.63)	(0.11)	(-0.65)	(-0.04)	(-0.21)	(0.35)	(-0.1)	(0.49)
D_PST * STK_DVF	-0.005 ***	-0.006 ***	-0.009 ***	-0.009 **	-0.012 **	-0.014 **	-0.016 *	-0.018 *
	(-2.81)	(-2.61)	(-2.87)	(-2.52)	(-2.24)	(-2.35)	(-1.88)	(-1.93)
R-sqr	*0.190*	*0.192*	*0.282*	*0.286*	*0.370*	*0.374*	*0.446*	*0.450*

*, **, *** Significant at the 0.10, 0.05, and 0.01 level, respectively. This table provides regressions of the change in the adverse selection components of Compustat firms on the change in the percentages of shares held by ETFs, with interaction terms with diversification dummies. I estimate pooled regressions of the firm sample and report coefficients and t-statistics calculated with two-way cluster robust errors (Gow et al. [2009]). **D_X**s refer to the change of variables calculated over one to four quarters of pre- and post-periods surrounding quarters when an ETF is newly introduced. Let t=0 be the event quarter and $X(t)$ denote a variable for a period that is t quarters before and after the event period. **D_X** is calculated as $\frac{1}{k}\sum_{t=1}^{k} X(t) - \frac{1}{k}\sum_{t=-k}^{-1} X(t)$, where k=1,…,4. For firms that are included in new ETFs, **ETF_DVF1** and **ETF_DVF2** are weighted by the number of shares held by these ETFs. **STK_DVF1 (STK_DVF2)** is 0 if the value-weighted **ETF_DVF1 (ETF_DVF2)** for the firm-quarter is closer to 0, and 1 if it is closer to 1. All variables are winsorized at 1%.

Table 7: Regressions of ETF Lambda on Holding Stocks' Lambdas

	Pred. Sign		Dep. Var: ETF_LAMBDA			
			ETF_DVF=			
	H2-a	H2-b	ETF_DVF1	ETF_DVF2	ETF_DVF1	ETF_DVF2
			1	2	3	4
Intercept			0.138 ***	0.126 ***	0.144 ***	0.132 ***
			(4.28)	(4.19)	(5.12)	(4.92)
ETF_DVF	**-**	**+**	**-0.068 ****	**-0.058 ***	**-0.125 *****	**-0.122 *****
			(-2.03)	**(-1.81)**	**(-3.18)**	**(-3.07)**
STK_LAMBDA			-0.076	-0.074		
			(-0.29)	(-0.29)		
STK_LAMBDA * ETF_DVF	**-**	**+**	**0.781 ***	**0.867 ****		
			(1.96)	**(2.22)**		
STK_ETF_%HLD					-0.363	-0.353
					(-0.29)	(-0.29)
STK_ETF_%HLD * ETF_DVF	**-**	**+**			**3.735 ***	**4.148 ****
					(1.96)	**(2.22)**
ETF_SIZE			-0.009 ***	-0.008 ***	-0.009 ***	-0.008 ***
			(-4.64)	(-4.5)	(-4.73)	(-4.56)
ETF_VOLUME			-0.003	-0.002	-0.003 *	-0.002
			(-1.34)	(-1.17)	(-1.94)	(-1.53)
ETF_STDRET			0.041 ***	0.040 ***	0.052	0.051
			(4.37)	(4.39)	(1.23)	(1.23)
ETF_ABS_ABNRET			0.124 **	0.108 *	0.131 ***	0.116 ***
			(2.08)	(1.87)	(3.31)	(2.95)
ETF_PRICE			-0.009	-0.008	-0.009	-0.009
			(-0.98)	(-1.06)	(-1.04)	(-1.15)
ETF_INST_%			0.017	0.017	0.018	0.017
			(1.49)	(1.46)	(1.61)	(1.58)
ETF_INST_N			-0.003	-0.004 *	-0.003	-0.003
			(-1.09)	(-1.84)	(-0.8)	(-1.3)
ETF_PST			-0.011	-0.014	-0.011	-0.014
			(-0.69)	(-0.91)	(-0.69)	(-0.91)
ETF_LAMBDA_MKT			0.420 ***	0.459 ***	0.408 ***	0.447 ***
			(6)	(6.37)	(6.67)	(6.93)
ETF_SIZE * ETF_DVF			0.007 ***	0.005 ***	0.006 ***	0.005 ***
			(3.62)	(3.45)	(3.58)	(3.32)
ETF_VOLUME * ETF_DVF			0.002	0.001	0.005 **	0.005 **
			(0.84)	(0.65)	(2.07)	(2.09)
ETF_STDRET * ETF_DVF			-0.024	-0.027	-0.141 *	-0.156 **
			(-1.18)	(-1.49)	(-1.82)	(-2.12)
ETF_ABS_ABNRET * ETF_DVF			-0.180 **	-0.179 **	-0.259 **	-0.266 ***

	(-2.03)	(-2.32)	(-2.59)	(-3.1)
ETF_PRICE * ETF_DVF	0.003	0.002	0.005	0.004
	(0.3)	(0.26)	(0.48)	(0.49)
ETF_INST_% * ETF_DVF	-0.010	-0.011	-0.013	-0.015
	(-0.76)	(-0.93)	(-1.03)	(-1.28)
ETF_INST_N * ETF_DVF	-0.003	-0.002	-0.006	-0.005 *
	(-1.1)	(-1.07)	(-1.58)	(-1.75)
ETF_PST * ETF_DVF	-0.006	0.004	-0.006	0.004
	(-0.34)	(0.2)	(-0.33)	(0.21)
ETF_LAMBDA_MKT * ETF_DVF	0.044	-0.025	0.172	0.116
	(0.47)	(-0.29)	(1.4)	(1.05)
R-sqr	*0.185*	*0.185*	*0.182*	*0.182*

*, **, *** Significant at the 0.10, 0.05, and 0.01 level, respectively. This table provides regressions of ETF-level adverse selection costs on the value-weighted averages of holding stock lambdas and other determinants of the ETF-level lambdas. I estimate pooled 2-stage least squares regressions of the ETFs sample. I use the idiosyncratic volatility of firm returns as an instrument in the first stage and estimate the predictive value of **STK_LAMBDA.** I report the second stage coefficients and t-statistics calculated with two-way cluster robust errors (Gow et al. [2009]). **ETF_LAMBDA** (in %) is the adverse selection component (λ) of ETF bid-ask spreads estimated per month with the method by Madhavan et al. [1997] modified by Armstrong et al. [2009]. Observations with negative values or larger than 100(%) of ETF_LAMBDA are excluded from the sample. **STK_LAMBDA** is the value-weighted average of holding stocks' lambdas of ETFs. **ETF_SIZE** the log of one plus ETF market value, calculated with the sum of stocks' daily market values (CRSP *prc* * *shrout*). **ETF_VOLUME** it the turnover of an ETF calculated as the log of one plus the quarterly average of daily CRSP trading volume divided by one plus the number of shares outstanding. **ETF_STDRET** is the annualized time-series standard deviation of daily returns, calculated each quarter. **ETF_ABS_ABNRET** is the absolute abnormal cumulative returns in each quarter. **ETF_PRICE** is the log of one plus the quarterly average of daily CRSP prices of an ETF. **ETF_INST_%** is the quarterly % of institutional ownership of an ETF from Thomson Financial Spectrum. **ETF_INST_N** is the log of one plus the quarterly number of institutional owners of an ETF from Thomson Financial Spectrum. **STK_IDIO** is the value-weighted average of holding stocks' idiosyncratic volatilities. Idiosyncratic volatility is calculated as the quarterly standard deviation of residuals of daily cross-sectional Fama-French 3-factor regressions. At least 15 daily observations are required to calculate the standard deviation for each quarter. **ETF_DVF1** is 0 if an ETF is a sector ETF, and 1 otherwise. **ETF_DVF2** is 0(1) if an ETF is ranked in the lower (higher) 50% of the factor score, calculated with the diversification score, the number of holding stocks, and the average weights of holding stocks. The diversification score is 0 if an ETF is a sector ETF, 1 if an ETF duplicates an index based on size and growth, 2 if an ETF duplicates an index based either on size or growth, and 3 if an ETF duplicates a broad market index. **ETF_PST** is the % of small trades, calculated with intraday trade data of ETFs. **ETF_LAMBDA_MKT** is the average of **ETF_LAMBDA,** calculated with all sample ETFs. All variables are winsorized at 1%. All variables are winsorized at 1%.

Table 8: Regressions of % of Small Trades

Panel A: Regressions of PST on ETF_%HLD

	Pred. Sign	Dep. Var: PST					
		1	2	3	4	5	6
Intercept		1.555 ***	1.581 ***	1.419 ***	1.425 ***	1.425 ***	1.587 ***
		(17.37)	(15.44)	(16.63)	(14.45)	(14.44)	(16.18)
ETF_%HLD	-		**1.935 *****		**2.402 *****	**2.402 *****	
			(12.08)		**(14.24)**	**(14.2)**	
MFID_%HLD	-			**-5.525 *****	**-6.649 *****	**-6.650 *****	
				(-11.2)	**(-15.3)**	**(-15.21)**	
ID_%HLD	-						**0.862 *****
							(5.16)
MFAM_%HLD	-					0.002	-0.029
						(0.07)	(-1.06)
SIZE	+	-0.044 ***	-0.043 ***	-0.035 ***	-0.032 ***	-0.032 ***	-0.045 ***
		(-17.26)	(-16.54)	(-16.42)	(-14.56)	(-14.56)	(-17.85)
BTM	-	-0.035 ***	-0.032 ***	-0.032 ***	-0.027 ***	-0.027 ***	-0.035 ***
		(-6.18)	(-5.8)	(-5.56)	(-4.96)	(-4.78)	(-6.02)
VOLUME	+	0.010 ***	0.007 ***	0.008 ***	0.005 ***	0.005 ***	0.009 ***
		(6.09)	(5.45)	(4.9)	(3.64)	(3.59)	(6.16)
STDRET	-	0.002 ***	0.002 ***	0.002 ***	0.003 ***	0.003 ***	0.002 ***
		(2.91)	(3.85)	(3.22)	(4.58)	(4.57)	(3.37)
ABS_ABNRET	-	-0.082 ***	-0.080 ***	-0.084 ***	-0.081 ***	-0.081 ***	-0.081 ***
		(-8.16)	(-9)	(-8.05)	(-9.06)	(-9.07)	(-8.54)
PRICE	-	-0.163 ***	-0.166 ***	-0.167 ***	-0.173 ***	-0.173 ***	-0.163 ***
		(-32.51)	(-35.39)	(-31.94)	(-35.4)	(-35.32)	(-33.87)
ANALYST	+	0.009 **	0.009 **	0.012 ***	0.011 ***	0.011 ***	0.009 **
		(2.65)	(2.55)	(3.37)	(3.44)	(3.57)	(2.73)
INST_%	-	0.059 ***	0.023 **	0.066 ***	0.022 **	0.021 **	0.048 ***
		(5.61)	(2.27)	(6.62)	(2.44)	(2.33)	(4.74)
INST_N	+	0.001	-0.002	0.006 **	0.004 **	0.004 **	-0.002
		(0.25)	(-0.91)	(2.59)	(2.06)	(2.09)	(-0.66)
PST_MKT	+	0.326 ***	0.242 *	0.339 ***	0.238 *	0.238 *	0.286 **
		(3.18)	(1.98)	(3.54)	(2.01)	(2.01)	(2.49)
R-sqr		*0.630*	*0.638*	*0.640*	*0.653*	*0.653*	*0.632*

*, **, *** Significant at the 0.10, 0.05, and 0.01 level, respectively. This table provides regressions of the percentages of small trades on the percentages of shares held by ETFs. I estimate pooled regressions of the firm sample and report coefficients and t-statistics calculated with two-way cluster robust errors (Gow et al. [2009]). **PST** is the % of small trades (defined as trades smaller than $5,000), calculated with the TAQ intraday trade dataset. **PST** is calculated daily and averaged by quarter. **ETF_%HLD** is the percentage of shares held by any of the U.S. equity ETFs, calculated with Thomson holdings data (S12). **MFID_%HLD** is the percentage of shares held by U.S. equity index mutual funds, calculated with Thomson holdings data (S12). **MFAM_%HLD** is the percentage of shares held by actively managed U.S. equity mutual funds. All non-ETF, non-index, U.S. equity mutual funds in Thomson holdings data (S12) are categorized as actively managed U.S. equity mutual funds. **ETF_%HLD, MFID_%HLD**, and **MFAM_%HLD** are calculated daily and averaged by quarter. **SIZE** is the log of one plus market value (CRSP *prc* * *shrout*). **BTM** is the book-to-market assets ratio calculated as (Compustat *ATQ*/(*ATQ* -Compustat *CEQQ* + Market Value))) on the quarter end date. **VOLUME** is the turnover calculated as the log of one plus the quarterly average of daily CRSP trading volume divided by one plus the number of shares outstanding. **STDRET** is the annualized time-series standard deviation of daily returns (CRSP *ret*), calculated each quarter. **ABS_ABNRET** is the absolute abnormal cumulative returns in each quarter. **PRICE** is the log of one plus the quarterly average of daily CRSP prices. **ANALYST** is the log of one plus the quarterly average number of analysts following counted from I/B/E/S. **INST_%** is the quarterly % of institutional ownership from Thomson Financial Spectrum. **INST_N** is the log of one plus the quarterly number of institutional owners from Thomson Financial Spectrum. **PST_MKT** is the average **PST** of all stocks in the sample. All variables are winsorized at 1%.

Panel B: Regression of D_PST on D_ETF_%HLD

	Pred. Sign	Dep. Var: PST			
		1 QTR	2 QTR	3 QTR	4 QTR
		1	2	3	4
Intercept		0.012 ***	0.021 ***	0.025 ***	0.028 ***
		(3.11)	(3.72)	(3.12)	(3.15)
D_ETF_%HLD	**-**	**-0.810 ***	**-1.071 ****	**-0.845 ***	**-0.509**
		(-1.85)	**(-2.5)**	**(-1.88)**	**(-1.08)**
D_SIZE	+	0.076 ***	0.056 ***	0.041 **	0.031 **
		(4.02)	(2.94)	(2.45)	(2)
D_BTM	-	-0.087 ***	-0.081 ***	-0.064 ***	-0.060 ***
		(-3.8)	(-3.34)	(-3.05)	(-3.21)
D_VOLUME	+	-0.020 ***	-0.005	0.004	0.012
		(-2.84)	(-0.64)	(0.44)	(1.19)
D_STDRET	-	0.000	0.001 *	0.001	0.000
		(1.63)	(1.73)	(1.06)	(0.4)
D_ABS_ABNRET	-	-0.032 ***	-0.076 ***	-0.122 ***	-0.173 ***
		(-4.05)	(-3.92)	(-5.79)	(-8.06)
D_PRICE	-	-0.408 ***	-0.378 ***	-0.354 ***	-0.342 ***
		(-15.14)	(-14.74)	(-14.21)	(-14.29)
D_ANALYST	+	0.012 ***	0.011 ***	0.014 **	0.012 **
		(4.03)	(2.78)	(2.57)	(2.45)
D_INST_%	-	0.061 **	0.094 ***	0.097 ***	0.103 ***
		(2.51)	(3.04)	(3.86)	(4.42)
D_INST_N	+	0.017 *	0.001	-0.006	-0.009
		(1.74)	(0.1)	(-0.66)	(-0.98)
D_PST_MKT	+	0.262 *	0.330 *	0.503 **	0.555 ***
		(1.75)	(1.81)	(2.45)	(3.33)
R-sqr		*0.310*	*0.342*	*0.375*	*0.417*

*, **, *** Significant at the 0.10, 0.05, and 0.01 level, respectively. This table provides regressions of the change in the percentages of small trades on the change in the percentages of those firms' shares that are held by ETFs. I estimate pooled regressions of the firm sample and report coefficients and t-statistics calculated with two-way cluster robust errors (Gow et al. [2009]). **D_X**s refer to the change of variables calculated over one to four quarters of pre- and post-periods surrounding quarters when an ETF is newly introduced. Let *t*=0 be the event quarter and *X(t)* denote a variable for a period that is *t* quarters before and after the event period. **D_X** is calculated as $\frac{1}{k}\sum_{t=1}^{k} X(t) - \frac{1}{k}\sum_{t=-k}^{-1} X(t)$, where *k*=1,…,4. All variables are winsorized at 1%.

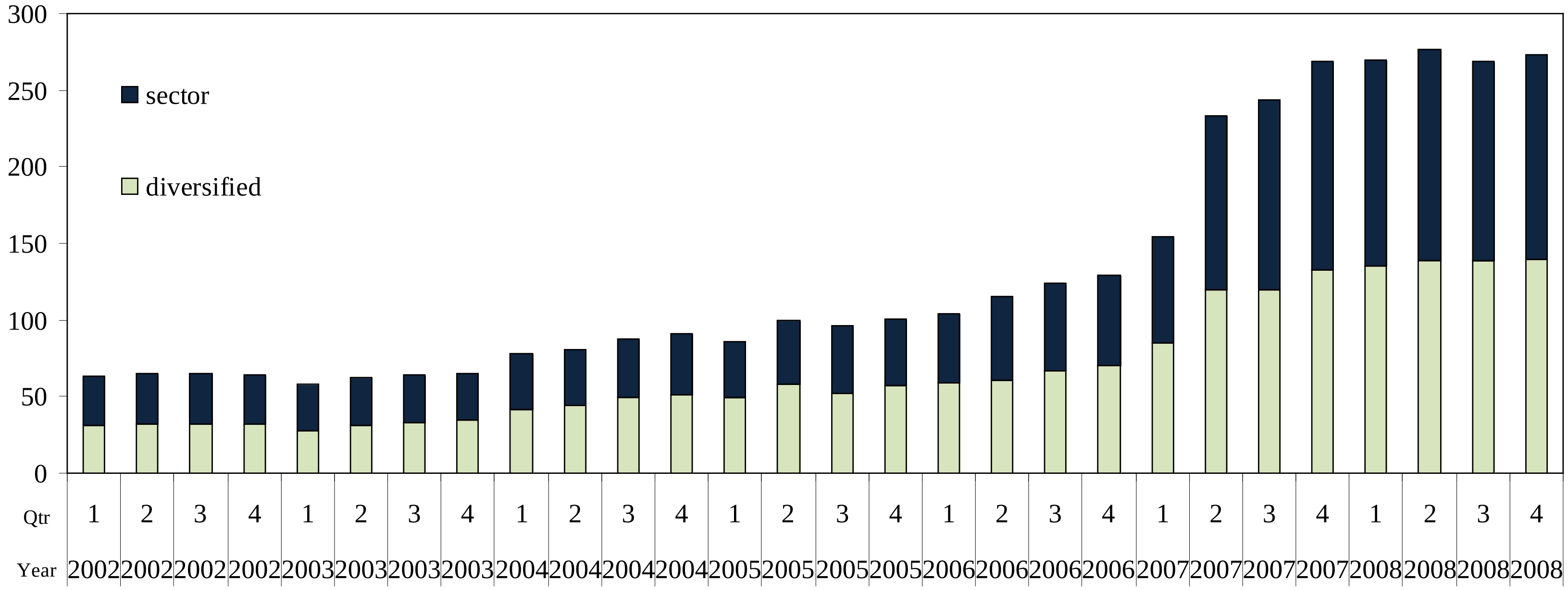
Panel C: Regressions of D_PST on diversification

		Dep. Var: D_PST							
		1 QTR		2 QTR		3 QTR		4QTR	
		STK_DVF=							
	Pred. Sign	STK_DVF1	STK_DVF2	STK_DVF1	STK_DVF2	STK_DVF1	STK_DVF2	STK_DVF1	STK_DVF2
		1	2	3	4	5	6	7	8
Intercept		0.011	0.013 *	0.027 **	0.019 ***	0.033 **	0.023 **	0.039 **	0.028 ***
		(1.26)	(1.91)	(2.34)	(2.72)	(2.13)	(2.5)	(2.49)	(3.12)
STK_DVF		0.001	0.000	-0.006	0.003	-0.008	0.003	-0.011	0.001
		(0.06)	(-0.04)	(-0.53)	(0.39)	(-0.49)	(0.29)	(-0.63)	(0.06)
D_ETF_%HLD		-0.113	-0.035	-0.589 *	-0.376 *	-0.711 **	-0.418 *	-0.681 *	-0.483 **
		(-0.38)	(-0.16)	(-1.91)	(-1.78)	(-2.08)	(-1.78)	(-1.92)	(-2.13)
D_ETF_%HLD* STK_DVF	**-**	**-0.939**	**-1.164 ****	**-0.695**	**-1.068 ***	**-0.227**	**-0.661**	**0.168**	**-0.127**
		(-1.62)	**(-2.47)**	**(-1.39)**	**(-1.88)**	**(-0.38)**	**(-0.98)**	**(0.3)**	**(-0.2)**
D_SIZE		0.084 **	0.057 *	0.064 **	0.057 **	0.053 *	0.045 *	0.046 *	0.033
		(2.28)	(1.92)	(2.11)	(2.5)	(1.88)	(1.93)	(1.82)	(1.47)
D_BTM		0.001	-0.003	0.032	0.012	0.055	0.026	0.066	0.019
		(0.01)	(-0.03)	(0.34)	(0.14)	(0.58)	(0.33)	(0.97)	(0.31)
D_VOLUME		-0.011	-0.012	-0.012	-0.006	0.009	0.009	0.025 **	0.022
		(-0.98)	(-1.2)	(-1.37)	(-0.66)	(0.7)	(0.75)	(2.12)	(1.62)
D_STDRET		0.000	0.000	-0.001	0.000	0.000	0.000	-0.002	-0.001
		(0.28)	(0.48)	(-0.53)	(-0.34)	(0.15)	(0.07)	(-0.87)	(-0.57)
D_ABS_ABNRET		0.003	-0.021	-0.014	-0.033	-0.103 ***	-0.116 ***	-0.131 ***	-0.157 ***
		(0.08)	(-0.81)	(-0.35)	(-0.83)	(-4.44)	(-3.29)	(-2.77)	(-4.49)
D_PRICE		-0.436 ***	-0.424 ***	-0.412 ***	-0.410 ***	-0.394 ***	-0.386 ***	-0.384 ***	-0.375 ***
		(-15.1)	(-16.25)	(-15.71)	(-16.38)	(-14.61)	(-15.98)	(-15.4)	(-16.37)
D_ANALYST		0.022 *	0.026 ***	0.026 **	0.025 **	0.028 **	0.028 ***	0.033 ***	0.033 ***
		(1.65)	(2.69)	(2.16)	(2.06)	(2.52)	(2.95)	(2.85)	(3.5)
D_INST_%		0.116 **	0.071 *	0.248 ***	0.182 ***	0.233 ***	0.171 ***	0.228 ***	0.178 ***
		(2.07)	(1.69)	(5.18)	(4.1)	(5.05)	(3.68)	(4.34)	(2.95)
D_INST_N		-0.077 ***	-0.059 ***	-0.060 ***	-0.055 ***	-0.044 ***	-0.046 ***	-0.045 ***	-0.048 ***
		(-3.66)	(-2.63)	(-4.33)	(-3.32)	(-2.75)	(-3.51)	(-2.76)	(-3.44)
D_PST_MKT		0.209	0.183 **	0.089	0.376 ***	0.193	0.512 ***	0.185	0.517 ***

	(1.57)	(2.4)	(0.48)	(3.29)	(0.82)	(3.53)	(0.93)	(3.45)
D_SIZE * STK_DVF	-0.011	0.023	-0.010	-0.001	-0.014	-0.005	-0.016	-0.002
	(-0.24)	(0.59)	(-0.28)	(-0.02)	(-0.43)	(-0.16)	(-0.56)	(-0.06)
D_BTM * STK_DVF	-0.095	-0.091	-0.123	-0.100	-0.132	-0.101	-0.139 *	-0.089
	(-0.84)	(-1.03)	(-1.22)	(-1.14)	(-1.27)	(-1.15)	(-1.85)	(-1.35)
D_VOLUME * STK_DVF	-0.011	-0.010	0.006	-0.001	-0.007	-0.007	-0.017	-0.015
	(-1.05)	(-0.92)	(0.83)	(-0.12)	(-0.55)	(-0.56)	(-1.46)	(-1.01)
D_STDRET * STK_DVF	0.000	0.000	0.002 *	0.002 **	0.001	0.001	0.002	0.001
	(0.17)	(-0.01)	(1.87)	(2.44)	(0.31)	(0.61)	(0.98)	(0.82)
D_ABS_ABNRET * STK_DVF	-0.038	-0.012	-0.066	-0.044	-0.019	-0.003	-0.043	-0.013
	(-1.08)	(-0.47)	(-1.61)	(-1.07)	(-0.54)	(-0.07)	(-0.74)	(-0.29)
D_PRICE * STK_DVF	0.034	0.022	0.040	0.040	0.047 *	0.041	0.049 *	0.041
	(0.97)	(0.72)	(1.39)	(1.41)	(1.7)	(1.45)	(1.87)	(1.51)
D_ANALYST * STK_DVF	-0.011	-0.015	-0.015	-0.015	-0.015	-0.015 *	-0.022 **	-0.023 ***
	(-0.78)	(-1.62)	(-1.26)	(-1.17)	(-1.42)	(-1.73)	(-2.06)	(-2.86)
D_INST_% * STK_DVF	-0.060	-0.011	-0.167 ***	-0.101 **	-0.151 ***	-0.086 *	-0.139 ***	-0.088
	(-0.91)	(-0.21)	(-2.73)	(-2.06)	(-3.18)	(-1.83)	(-2.65)	(-1.4)
D_INST_N * STK_DVF	0.104 ***	0.085 ***	0.067 ***	0.064 ***	0.042 **	0.045 ***	0.040 **	0.044 ***
	(5.83)	(3.59)	(3.59)	(3.6)	(2.21)	(3.11)	(2.06)	(2.81)
D_PST_MKT * STK_DVF	0.064	0.094	0.259	-0.059	0.333	-0.015	0.403 *	0.048
	(0.33)	(0.6)	(1.12)	(-0.28)	(1.16)	(-0.07)	(1.66)	(0.25)
R-sqr	*0.313*	*0.314*	*0.345*	*0.346*	*0.378*	*0.379*	*0.420*	*0.421*

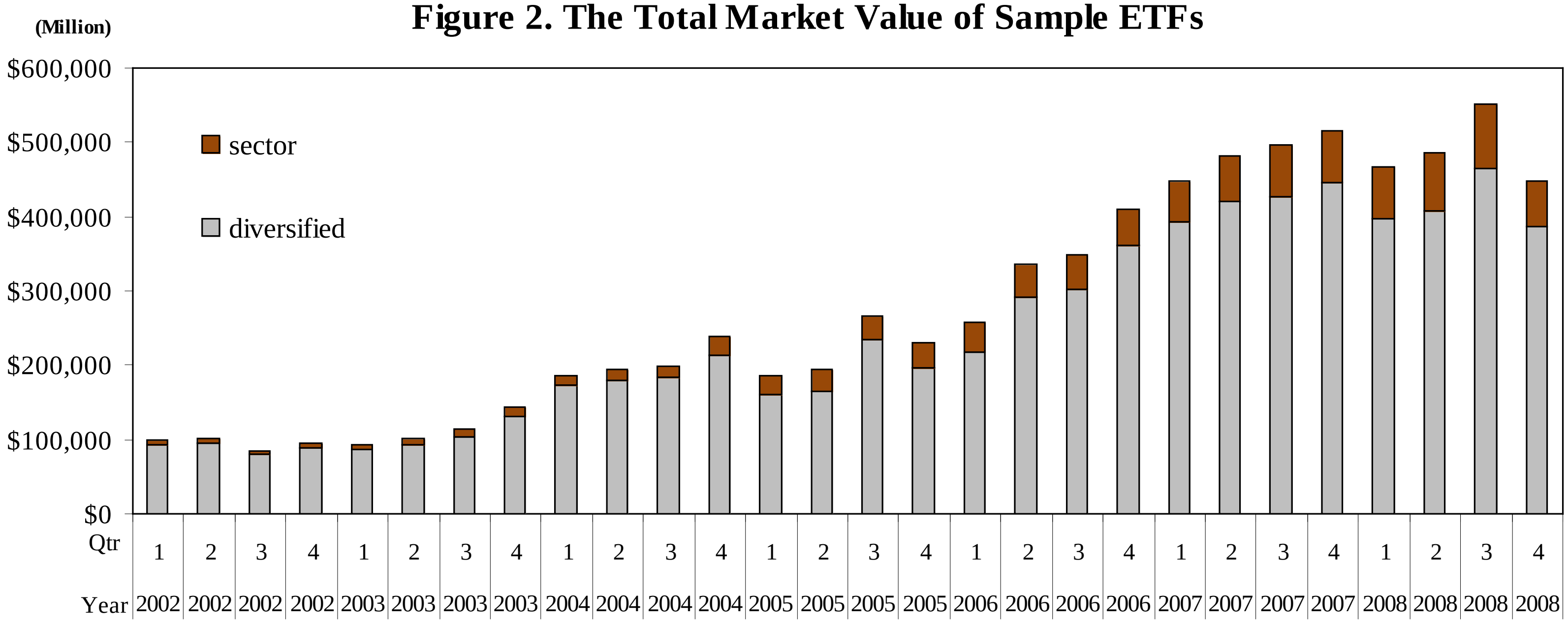
*, **, *** Significant at the 0.10, 0.05, and 0.01 level, respectively. This table provides regressions of the change in the percentages of small trades on the change in the percentages of shares held by ETFs, with interaction terms with diversification dummies. I estimate pooled regressions of the firm sample and report coefficients and t-statistics calculated with two-way cluster robust errors (Gow et al. [2009]). **D_X**s refer to the change of variables calculated over one to four quarters of pre- and post-periods surrounding quarters when an ETF is newly introduced. Let t=0 be the event quarter and $X(t)$ denote a variable for a period that is t quarters before and after the event period. **D_X** is calculated as $\frac{1}{k}\sum_{t=1}^{k} X(t) - \frac{1}{k}\sum_{t=-k}^{-1} X(t)$,where k=1,…,4. For firms that are included in new ETFs, **ETF_DVF1** and **ETF_DVF2** are weighted by the number of shares held by these ETFs. **STK_DVF1 (STK_DVF2)** is 0 if the value-weighted **ETF_DVF1 (ETF_DVF2)** for the firm-quarter is closer to 0, and 1 if it is closer to 1. All variables are winsorized at 1%.

Figure 1. The Number of Sample ETFs



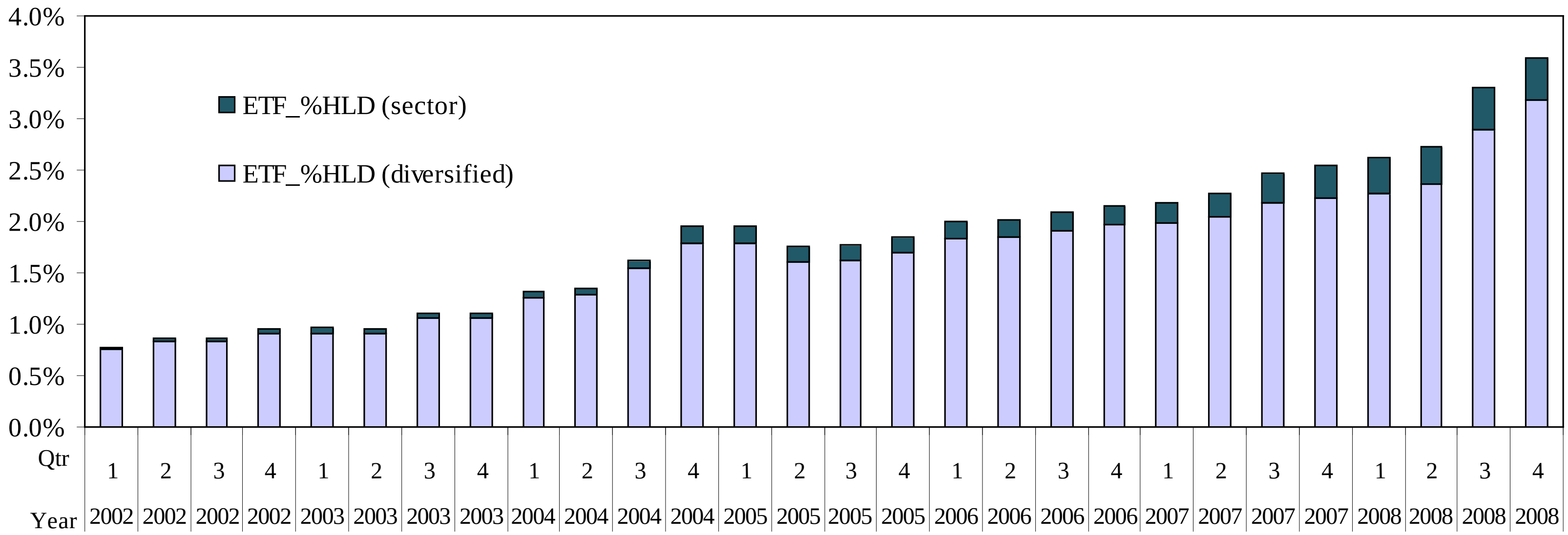
This figure presents the number of diversified ETFs (ETF_DVF1=1) and the number of sector ETFs (ETF_DVF0) in each quarter.

Figure 2. The Total Market Value of Sample ETFs



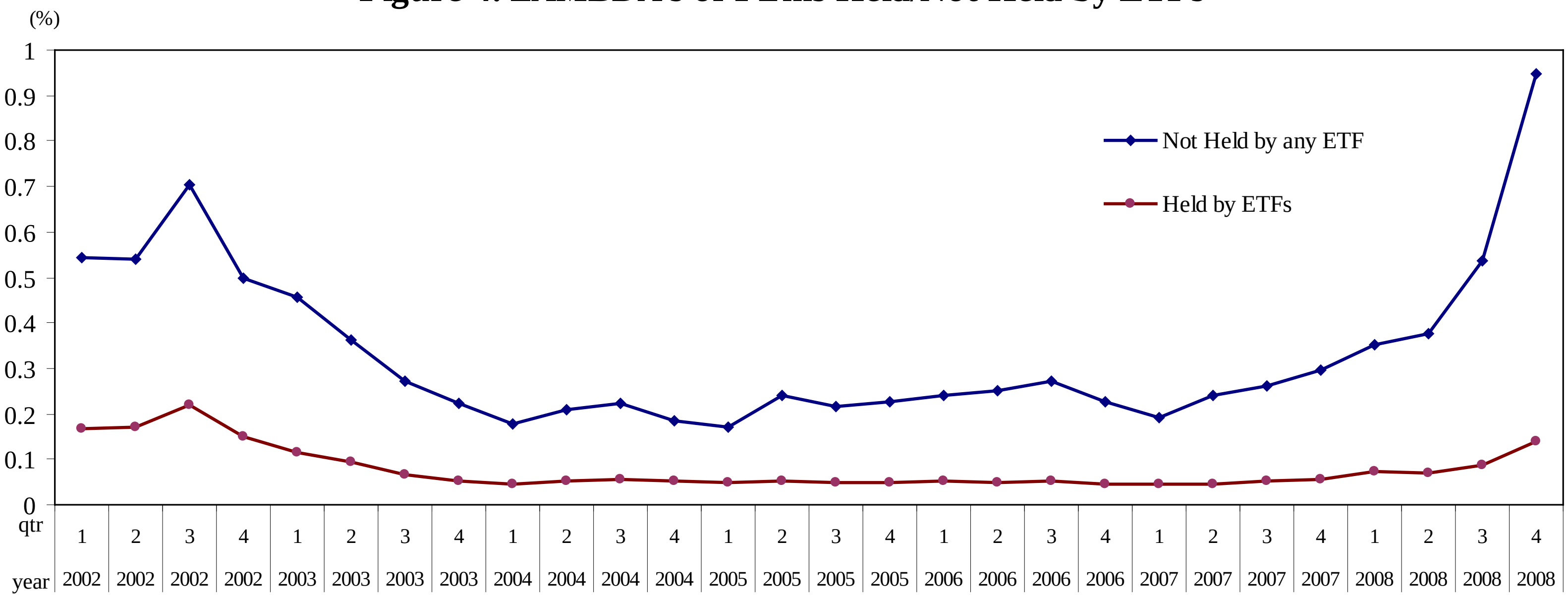
This figure presents the market value (in $million) of diversified ETFs (**ETF_DVF1**=1) and the market value (in $million) of sector ETFs (**ETF_DVF1**=0) in each quarter.

Figure 3. The Percentage of Firm Shares Held by ETFs (*ETF_%HLD*)



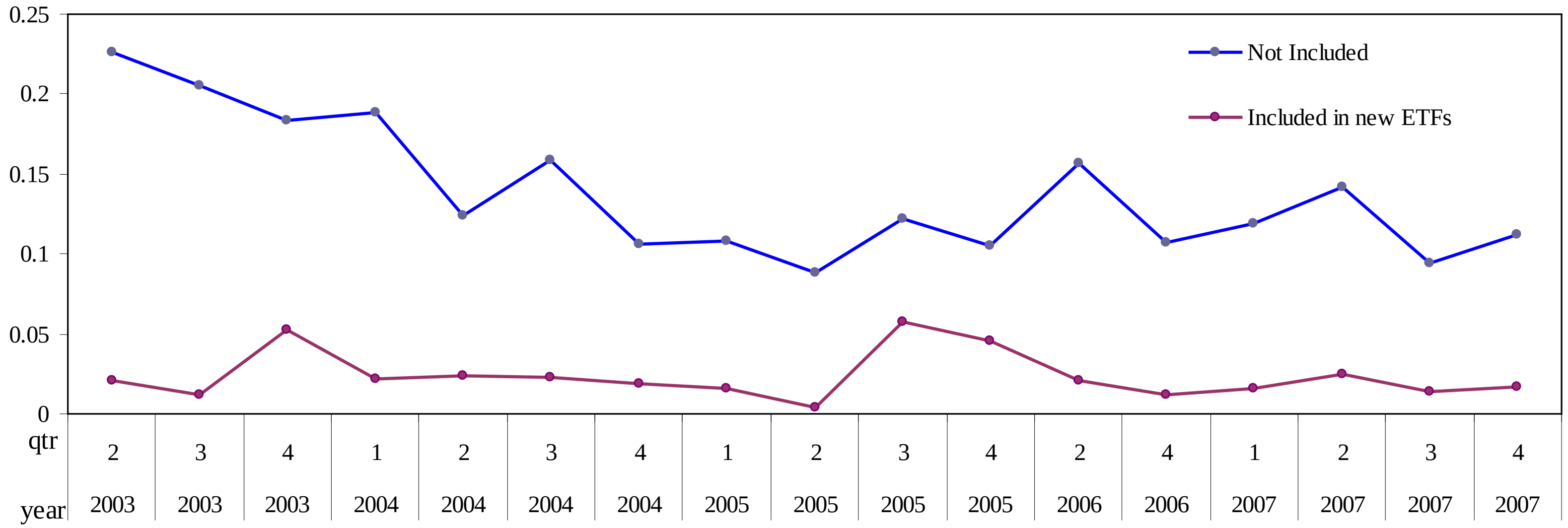
This figure presents the average of **ETF_%HLD** calculated with diversified ETFs (**ETF_DVF1**=1) and the average of **ETF_%HLD** calculated with diversified ETFs sector ETFs (**ETF_DVF1**=0) in each quarter.

Figure 4. *LAMBDA* s of Firms Held/Not Held by ETFs



This figure presents the average **LAMBDA** of firms that are not held by any ETF (**ETF_%HLD**=0) and the average **LAMBDA** of firms that are held by ETFs (**ETF_%HLD**>0) in each quarter.

Figure 5. *LAMBDA* s of Firms Included/Not Included in New ETFs Prior to the Events of New ETF Inception



This figure presents the average **PRE_LAMBDA** of firm-quarter observations that are not in the event study sample and the average **PRE_LAMBDA** of firm-quarter observations that in the event study sample (due to inclusion in the newly introduced ETFs). **PRE_LAMBDA** is **LAMBDA**, one quarter prior to the event quarter, calculated as $\frac{1}{k}\sum_{t=-k}^{-1} LAMBDA(t)$, where k=1. **PRE_LAMBDA**s calculated with k=2,…4 yield similar figures.