Are ETFs Replacing Index Mutual Funds?

Ilan Guedj and Jennifer Huang*

November 2008

Abstract

Flows to an Open-Ended Mutual Fund (OEF) can significantly hamper its subsequent performance due to flow-induced trading costs. An Exchange-Traded Fund (ETF) is designed not to have this cost and hence is advertised as the more efficient index vehicle. We develop an equilibrium model to investigate whether the ETF structure is indeed the dominant organizational form. We find that while flow-induced trading is costly to all OEF investors, it is beneficial to those who cause the flow – it is simply a zero-sum game. The OEF structure can therefore be viewed as providing partial insurance against future liquidity needs and is ex-ante beneficial for risk averse investors. However, the insurance feature embedded in the OEF structure can cause moral hazard issues – e.g., excessive trading – and reduce the performance of the OEF. We also find that investors with higher liquidity needs benefit more from the liquidity insurance and hence prefer to invest via the OEF. Interestingly, the concentration of high-liquidity-need investors in the OEF does not lead to higher flow-induced trading costs since individual liquidity needs cancel out at the fund level. As a result, OEFs are not dominated by ETFs in equilibrium. Finally, guided by the theoretical prediction that ETFs are better suited for narrower and less liquid underlying indexes, we empirically confirm that the growth of ETFs is more concentrated in those indexes.

*McCombs School of Business, University of Texas at Austin. Guedj: `guedj@mail.utexas.edu` and (512) 471-5781. Huang: `jennifer.huang@mccombs.utexas.edu` and (512) 232-9375. The authors thank Franklin Allen, Tarun Chordia, Douglas Diamond, Darrell Duffie, William Goetzmann, David Musto (discussant), Francisco Perez-Gonzalez, Oleg Rytchkov, Jacob Sagi, Clemens Sialm, Paolo Sodini, Johan Sulaeman (discussant), Sheridan Titman, Charles Trzcinka (discussant), Luis Viceira, participants at the Institutional Investors Conference at UT-Austin, the Lone Star Conference, and 19th Annual Conference on Financial Economics and Accounting, and seminars at BI Norwegian School of Management and Stockholm School of Economics for comments and suggestions. Support from Q-group is gratefully acknowledged.

1 Introduction

Since the introduction of the first Exchange-Traded Fund (ETF) in the U.S. in 1993, ETFs have captured most of the growth in index mutual funds and now constitute about 40% of the index fund market share. Despite their growth, not all practitioners are convinced of their value as indexing tools. On the one hand, John Bogle, founder of Vanguard, has been the most vocal critic of ETFs: *"if long-term investing was the paradigm for the classic index fund, trading ETFs can only be described as short-term speculation."*[1] On the other hand, Lee Kranefuss, CEO of iShares at Barclays Global Investors, argues that ETFs are to mutual funds what compact discs were to records: *"a better product with more features for less money,"* especially since *"long-term ETF investors don't subsidize the costs of active traders in ETFs."*[2]

The view that ETFs are a more efficient indexing vehicle is rooted in the fact that fund flows to an OEF can be costly. In particular, whenever OEF investors purchase or redeem shares, their demand is pooled at the fund level and transacted at the daily closing price of the fund, which does not fully account for the future price impact that the fund may experience in implementing the pooled demand. Thus, there is cross-subsidization among investors – either existing investors subsidize new investors in the case of net fund inflows or the remaining investors subsidize the departing investors in the case of net fund outflows. This is the cost of the OEF structure that Kranefuss and other ETF advocates have emphasized. This cost is also well recognized in the academic literature. For example, Edelen (1999) documents that the extra trading (and the resulting price impact) induced by fund flows can reduce the average performance of a fund by as much as 1.4% per year.[3] Greene and Hodges (2002) find that even daily fund flows, which are more transient and less likely to cause mutual fund managers to incur transaction costs, can impose significant costs on the fund, especially for funds with large daily flows. Dickson, Shoven, and Sialm (2000) demonstrate that shareholder flows negatively affect mutual funds' after-tax returns. Johnson (2004) shows that different types of investors impose different flow costs on the fund, and Christoffersen, Keim, and Musto (2007) find that index funds have even *higher*

[1] "'Value' Strategies", WSJ, John Bogle, February 9, 2007.

[2] See Viceira and Wagonfeld (2007) for a detailed description of BGI's success in the ETF industry.

[3] Gastineau (2004) applies Edelen (1999)'s cost estimation and concludes that the cost of providing liquidity for open-ended mutual funds can be as high as $40 billion per year.

trading costs than active funds – despite the lack of information content of their trades – due to the inflexibility of index funds in meeting liquidity needs.[4] Overall, this literature points to significant costs of the OEF structure associated with flow-induced trading.[5]

ETFs, in contrast, are designed not to have flow-induced trading costs. Like closed-ended index mutual funds, ETFs are traded on the stock exchange with individual traders bearing their own transaction costs. In addition, ETF shares can be created or redeemed using the underlying index assets (in-kind redemption). This feature is often used to arbitrage away any price deviation between the ETF and the underlying stocks. By avoiding the need to purchase or liquidate the underlying assets, ETFs do not incur any trading-related costs upon the creation or redemption of shares.

In this paper, we develop an equilibrium model to investigate whether the ETF structure is indeed the dominant organizational form relative to an Open-Ended Index Mutual Fund (OEF). We compare their relative advantages in meeting liquidity demand for small investors and ask whether the rapid growth of the ETF industry implies a permanent structural shift in the mutual fund industry.

Our first result is that ETFs are not more efficient than OEFs – it is a zero-sum game between those who cause the fund flow and those who bear the flow-induced trading costs. Indeed, the OEF structure can be viewed as providing partial insurance against future liquidity needs: Investors pay the cost of lower average returns in exchange for better prices when they experience liquidity shocks.[6] As long as investors are risk averse, the OEF structure is actually beneficial rather than costly to investors. However, this benefit may not be obvious as the reported performance of OEFs can be lower than that of ETFs due to a dif-

---

[4]Grinblatt and Titman (1989) are the first to point out the significant transaction costs incurred by actively managed funds and show that the total costs can be as high as 2.5% per year. Edelen, Evans, and Kadlec (2007) find that the annual trading costs for a large sample of equity funds are comparable in magnitude to their expense ratio. These costs are higher for larger funds, especially if the fund has larger relative trade sizes, suggesting that trading costs contribute to the diseconomies of scale for mutual funds identified in Chen, Hong, Huang, and Kubik (2004).

[5]Sophisticated investors may even exploit the pricing scheme of mutual funds to design profitable trading strategies. For example, Chalmers, Edelen, and Kadlec (2001) find that on extreme days the failure to account for nonsynchronous trading in determining funds' net asset value can result in an average one-day excess return of 0.84 percent at high beta small-cap domestic equity funds. Goetzmann, Ivkovic, and Rouwenhorst (2001) show that the issue of nonsynchronous trading is most pronounced in international mutual funds and propose a "fair pricing" mechanism that partially corrects net asset values for stale prices. Zitzewitz (2006) provides evidence that some investors may even abuse this pricing feature by trading after 4pm, and that such late trading can lead to significant shareholder welfare loss.

[6]This is similar to the insurance feature of bank deposit contracts in Diamond and Dybvig (1983).

ference in the way returns are accounted for. The flow-induced transaction costs are equally shared by all investors in the OEF and reduce reported OEF performance, whereas they are incurred only by those ETF investors with liquidity needs and do not affect reported ETF performance.

Second, we find that the OEF structure is not without some downside. In particular, OEFs can cause moral hazard issues that induce excessive trading and increase the cost of the liquidity insurance. Since OEF investors transact at a price that is unaffected by their own trading decisions, there is little incentive for them to internalize the price impact of their orders. As a result, they tend to trade too aggressively given their trading needs, leading to excess impact on the equilibrium asset prices. This extra price impact makes the liquidity insurance an ex-ante negative-sum game and leads to a lower average return for the OEFs.

Third, the tradeoff between the liquidity-insurance benefit and the moral hazard cost differs for investors with different individual liquidity needs. Quite intuitively, high-liquidity-need investors benefit more from the liquidity insurance and hence prefer to invest via the OEFs, whereas low-liquidity-need investors find the excess price impact induced by moral hazard too costly and hence prefer to invest via the ETFs. This result runs counter to John Bogle's critique that "ETFs can only be described as short-term speculation," and suggests a viable role for ETFs in the mutual fund industry.

Fourth, we show that the concentration of high-liquidity-need investors in the OEF does not lead to higher flow-induced trading costs, since individual liquidity needs cancel out at the fund level. As a result, the OEF structure is also viable in equilibrium. Moreover, as long as high- and low-liquidity-need investors have similar exposure to the systematic liquidity risk, they contribute equally to the trading costs of the OEF.[7] This result highlights a potential pitfall of the recent trend in the mutual fund industry to impose frequent-trading restrictions: higher frequency traders do not pose much higher costs to the OEF yet they benefit much more from the liquidity insurance provided by the OEF structure. If they are deterred from investing in the OEFs by trading restrictions, then OEFs can potentially lose a significant client base.

Finally, our model provides a framework to assess the effectiveness of the OEF and the

[7]We focus on the flow-induced trading costs for the OEFs and abstract away other operational costs for OEFs, like order execution, book keeping and so on. Since all these other costs have a large fixed component and decrease with the fund scale while the trading costs increase with the fund scale, our model is better suited for funds of reasonably large size.

ETF structures under different circumstances and to make predictions regarding the long-term trend in the growth of the mutual fund industry. One prediction of the model is that if investors have more correlated liquidity needs, the OEF is expected to have larger unbalanced demand from its investors. The price impact is higher, making the OEF less attractive relative to the ETF as an indexing vehicle. As a result, we expect to see a smaller size of the OEF industry in equilibrium. This prediction is empirically testable. For example, it is reasonable to assume that investors in a narrower index (such as a bio-tech index) are more likely to have correlated liquidity needs compared to investors in a broader market index (such as the S&P 500 index). Hence, we expect to have more ETFs in narrower indexes than in broader market indexes. Similarly, less liquid underlying indexes are likely to generate a higher price impact and larger tracking error for OEFs, making them less ideal candidates for OEF investing as well.

Empirically we perform several analyses. First, we document the growth pattern of ETFs and OEFs, both in broader market indexes and in specialty funds. Second, based on our theoretical predictions, we investigate whether the predicted cross-sectional differences between OEFs and ETFs exist in the data. We find that ETFs have grown not only in terms of total net assets under management, but also in terms of number of indexes tracked. Although there are approximately the same number of ETFs and OEFs available in the marketplace, ETFs track more than four times as many indexes as OEFs do. Moreover, we document that the indexes tracked vary significantly both among ETFs and across ETFs and OEFs. As predicted by our theoretical model, there are cross-sectional differences between OEFs and ETFs. This explains part of the observed growth of the ETF industry: Many new ETFs track indexes that have very high volatility, low liquidity, or high industry concentration that would have a high tracking error were they in an OEF, and hence are better off being in an ETF. This also implies that one needs to be careful when extrapolating the overall growth rate of ETFs, as this rate may exaggerate the ETF growth in some sectors of the market.

Despite the phenomenal growth of the ETF industry, there is limited academic research on ETFs, mostly due to this sector's short history. Elton, Gruber, Comer, and Li (2002) study SPDR, the first ETF that tracks the S&P 500 index, and document that it underperforms relative to both the underlying index and to other OEFs tracking the same index. Poterba and Shoven (2002) look at the tax implications of ETFs in general and the performance of SPDR in particular. They find that although in theory ETFs can be more tax efficient, in reality SPDR ETF performs slightly worse than the Vanguard S&P 500 both in before-tax and after-tax returns. We add to this literature by placing ETFs in a broader context, to understand their impact on the structure of the mutual fund industry. Agapova (2006) compares fund flows into index mutual funds and ETFs and finds that their coexistence can be partially explained by a clientele effect.

Our paper is most related to Chordia (1996) and Chen, Goldstein, and Jiang (2007), who consider the cross-subsidization induced by the pricing scheme of the OEF structure. Chordia (1996) is the first to point out that this cross-subsidization can be viewed as a form of insurance against liquidity shocks along the line of Diamond and Dybvig (1983). His focus, however, is quite different. He shows that funds optimally hold more cash in order to meet redemption demands and that load and redemption fees can help alleviate liquidity impacts. Chen, Goldstein, and Jiang (2007) show that the cost of this cross-subsidization can manifest into a "bank run" in which investors flee the fund for fear of bearing the liquidity cost imposed by others' withdrawing from the fund. We show that both features are present: While the liquidity insurance feature embedded in OEFs is beneficial, especially for higher-liquidity-need investors, it can induce moral hazard issues and reduce the performance of ETFs. As a result, lower-liquidity-need investors may prefer the ETF structure.

Another related literature investigates the interaction between investor behavior and the organizational structure of mutual funds. Nanda, Narayanan, and Warther (2000) show that the existence of investor clienteles with differing liquidity and marketing needs gives rise to a variety of open-ended fund structures that differ in the average return delivered to investors. Massa (1997) suggests that the vast number of funds offered to investors can be seen as marketing strategies used to exploit investor heterogeneity and that market forces may induce a sub-optimal number of mutual funds and categories. Christoffersen and Musto (2002) show that the price sensitivity of individual investors affects the pricing scheme of mutual funds. Our paper complements this literature by introducing a new investment vehicle (an ETF) with a different trading mechanism and studying investor choice between the new and the existing vehicles in equilibrium.

This paper also relates to the literature on the structure of the mutual fund industry. The efficiency of the open-ended structure has been much debated. As discussed earlier, a large literature documents the cost of the OEF structure associated with flow-induced trading. Separately, a smaller, mainly theoretical literature, suggests a potential benefit of the OEF structure when managers have ability. For example, Berk and Green (2004) show that when investors vote with their feet in open-end mutual funds, they can efficiently align money flow with managerial talent. Stein (2005) further argues that when open-ending is the only creditable signal of managerial ability, it becomes the dominant structure in the mutual fund industry despite its apparent inefficiency in allowing skilled managers to take advantage of arbitrage opportunities (e.g., Shleifer and Vishny (1997)). Combining the two literatures, one is tempted to conclude that the OEF structure is costly in meeting liquidity needs while it might be necessary for rewarding superior abilities. We add to the debate by showing that, even in the absence of ability, the OEF structure has its merits relative to the ETF structure. On the practical front, our analysis suggests that the rush in the industry to develop active ETFs – in addition to presenting the technical difficulty of revealing the portfolio holding for in-kind redemption while maintaining anonymity – may be ill motivated.

Finally, we contribute to the literature on Closed-Ended Mutual Funds (CEF). While the earlier literature largely relates the discount/premium of CEFs to irrational behavior of investors and the cost of arbitrage (see, for example, Lee, Shleifer, and Thaler (1991) and Pontiff (1996)), recently more effort has been devoted to explaining the discount/premium using rational models with managerial ability and liquidity demands (see, for example, Chay and Trzcinka (1999), Berk and Stanton (2007) and Cherkes, Sagi, and Stanton (2007)). Viewing ETFs as closed-ended index funds, we establish a benchmark at which neither managerial ability nor a discount/premium exists. Our framework, which captures the main difference between the OEF and ETF structures, can provide a starting point for understanding the difference between the ETF and CEF structures and in turn shed light on the phenomenal growth of ETFs and the stagnation of CEFs (which are active ETFs except for the in-kind redemption feature).

The remainder of the paper proceeds as follows. Section 2 describes the ETF industry and how it has evolved. Section 3 sets up the model and Section 4 describes the equilibrium. Section 5 discusses the properties of the equilibrium and derives empirical implications of the model. Based on these implications, Section 6 describes our empirical investigation. Section 7 concludes.

2 The growth of the ETF industry

2.1 Differences between ETFs and OEFs

ETFs are closed-ended investment companies that can be traded at any time throughout the course of the day on the secondary market. ETFs try to replicate stock market indexes. Contrary to regular index OEFs, ETFs are required to hold all the stocks that comprise the index they track.

The first ETF was introduced on the Toronto Stock Exchange in 1990. The first U.S. ETF was introduced in 1993 by State Street Global Advisors – SPDR (Standard & Poor's Depository Receipts, or "Spiders"), which tracks the S&P 500 index – and it is still the largest ETF by market cap. Most stock exchanges around the world now offer ETFs.

ETFs are closed-ended and not open-ended mutual funds and hence have two main differences from regular open-ended mutual funds. First, they are traded on the secondary market and therefore can be traded during the day and not only once a day. Second, if there is an inflow of money to the fund, the fund does not directly create new shares. However, contrary to standard closed-ended mutual funds, new shares can be created (or redeemed). The creation and redemption of shares in ETFs is also different from that of OEFs. Rather than the fund manager dealing directly with shareholders, certain parties, such as institutional investors, who have entered into a contract with the fund (called Authorized Participants (APs)) will create a basket of shares replicating the index that the fund tracks, and deliver them to the fund in exchange for ETF shares. A basket, or creation unit, consists of anywhere from 10,000 ETF shares to 600,000 ETF shares. ETF shares are then sold and resold freely among investors on the open market. If an investor accumulates a sufficient amount of ETF shares, the investor can exchange one full creation unit of ETF shares for a basket of the underlying shares of stock. The ETF creation unit is then redeemed, and the underlying stocks are delivered out of the fund. One of the advantages of this creation/redemption process for the fund investors is that institutional investors cover the dealing costs in purchasing the required shares to make up the portfolio. This mechanism allows institutional investors to take advantage of arbitrage opportunities when the price of the ETF deviates from its Net Asset Value (NAV). It is also the main distinction between an ETF and a regular closed-ended mutual fund, which results in a very small discount/premium between the price and the NAV. Thus, none of the regular discount/premium characteristics discussed in the literature with regard to closed-ended mutual funds apply in any substantial way to ETFs.

ETFs and OEFs also have different tax advantages. Whenever an OEF realizes a capital gain that is not balanced by a realized loss, the mutual fund must distribute the capital gains to shareholders by the end of the quarter. This can happen when stocks are added to and removed from the index, or when a large number of shares are redeemed. These gains are taxable to all shareholders. In contrast, ETFs are not redeemed by holders (instead, holders simply sell their ETFs on the stock market, as they would a stock), so investors generally only realize capital gains when they sell their own shares. However, there are some potential taxation drawbacks to ETFs. First, ETFs have to hold the exact mirror image of an index, while OEFs do not. Hence, around changes in the composition of an index, the ETF may have to sell existing stocks in order to re-balance its holdings. OEFs do not have to hold the mirror image of the index, and hence have more flexibility around index changes. Second, ETFs often trade their shares more rapidly to maintain a high cost basis of their underlying shares, which can result in ETF dividends failing to be classified as qualified dividends since the underlying shares don't satisfy IRS requirements. This can be a substantial drawback since one's ordinary tax rate may be significantly higher than the 15% tax charged on qualified dividends.

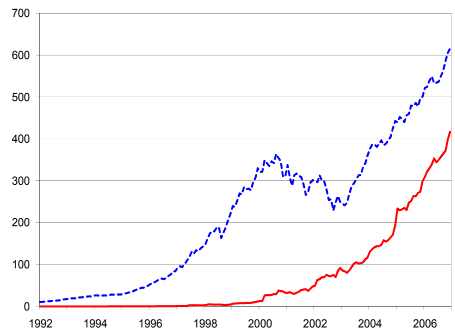
Note that ETFs are structured either as a mutual fund or as a Unit Investment Trust (UIT). A UIT is a U.S. investment company offering a fixed portfolio of securities that have a finite life. UITs are assembled by a sponsor and sold through brokers to investors.

2.2 Trends in the ETF industry

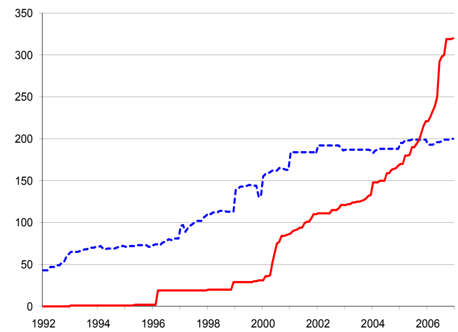
As we discuss in the introduction, while ETFs were introduced in the U.S. only in 1993, they have since grown at an exponential rate from one ETF to more than 320 in 2006.

Figure 1(a) shows the significant increase in money invested in ETFs between the early 1990s and the present. As one can see, the growth in ETFs was clearly slow between 1993 and 2000. This is for two main reasons. First, ETFs were a new investment vehicle, and as such were not yet very familiar to investors. Second, there were few ETFs available, making it a relatively exotic investment vehicle. Figure 1(b) depicts the relative number of ETFs available compared to OEFs over time. By year 2000, there were only about 20 ETFs available covering the major indexes, while there were more than 150 OEFs available. A direct comparison of the number of available ETFs to the number of available OEFs,

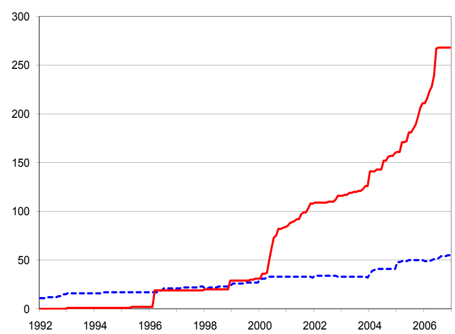
(a) Dollar Amount Invested



(b) Number of Funds



(c) Number of Indexes Tracked



(d) Average Total Net Assets

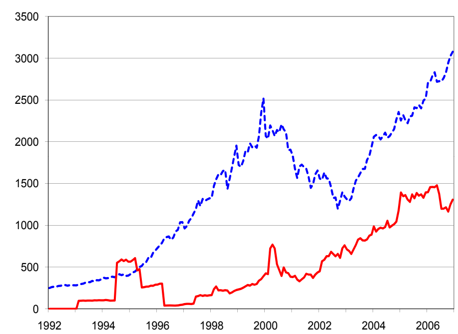


Figure 1: **Summary statistics of the ETF and the OEF industries.** In all panels, the dotted and the solid lines refer to OEFs and ETFs, respectively. Panel (a) reports the dollar amount invested (in billion dollars), panel (b) reports the number of funds, panel (c) reports the number of indexes tracked, and panel (d) reports the average total net assets of the funds (in million dollars).

however, is slightly misleading. In the early years each new ETF tracked a different index; even now, there are few ETFs that track exactly the same index. OEFs, in contrast, work somewhat differently. Mutual funds depend on their respective distribution channels, and not all mutual funds are available to all investors. For example, an investor with a Vanguard account can only invest in Vanguard mutual funds. Investors in retirement accounts follow even stricter rules and have limited offerings that depend on the company administering the account. The result is that in the past 40 years each mutual fund company (or at least the larger ones with their own distribution channels) has come to offer funds that are very similar to those of other mutual fund families. In the case of index mutual funds many families offer essentially identical funds that track the same index. For example, there are more than a dozen OEFs that track the S&P 500, while not long ago there was only one ETF. Even now there are only two ETFs tracking this index. Thus, while there are more OEFs than ETFs until very recently, OEFs are tracking fewer indexes.

Figure 1(c) shows the large increase in the number of different indexes tracked by ETFs in the past seven years. Comparing Figure 1(c) to 1(a), we can see that part of the success of ETFs and the apparent market share they seem to have captured from OEFs is correlated with the huge increase in the number of indexes that are being tracked. This connection is related to our theoretical argument that the characteristics of the underlying indexes tracked may be related to the increased interest in these indexes. As Figure 1(a) shows, in the past 10 years, while ETFs have grown at a faster pace than OEFs, increasing their relative importance in the marketplace, they have also offered new indexes as investment vehicles at an exponential rate. Together, these figures imply that, indeed, there could be more to the rise in ETFs than the mere fact that they offer an advantage to small investors compared to traditional index OEFs. In Section 6 we attempt to analyze the data available to date to better understand the phenomenal rise of this new investment vehicle and, specifically, its long-run implications for the market and on the mutual fund industry.

Though this impressive growth of ETFs can be partially attributed to the diversity of offerings in terms of the number of indexes covered, Figure 1(d) indicates that the average ETF has comparable size to the average OEF. Admittedly, since there are many OEFs tracking each index and the decision of small investors to invest in them depends on their respective distribution channels, this is far from conclusive. It does indicate, however, that the growth process of this relatively new industry is a compilation of several simultaneous trends: The intrinsic growth of existing ETFs, the introduction of new ETFs covering existing indexes, and the offering of ETFs covering new indexes that are not offered by traditional OEFs. This evidence highlights the importance of the question at the root of this paper, namely, whether the apparent increase in the size of the ETF industry is going to eventually alter the landscape of the delegated portfolio management industry and, in particular, of the existing mutual fund structure.

3 Model

We construct a parsimonious model that captures the main difference between OEFs and ETFs: Orders are cleared in the ETF market directly and traders bear their own trading costs, whereas orders are pooled in the OEF before being submitted to the underlying stock market and all investors in the OEF (whether or not they trade) equally share any resulting trading costs. We model agents' trading needs by introducing heterogeneity in their risk exposures, which motivates them to trade for risk sharing.

3.1 Securities market

There are three dates, $t = 0, 1$, and 2. A risky stock S is traded in a competitive market. The stock yields a final payoff of V at time 2, which is normally distributed,

$$V \sim N(\bar{V}, \sigma_v^2). \tag{1}$$

The per capita supply of the stock is $\bar{\theta}$. In addition, there is a short-term riskless security, which yields a constant interest rate of zero.

In addition to trading the stock directly, investors can participate in either an OEF (denoted by F) or an ETF (denoted by E). Our main objective is to compare the effectiveness of the OEF and the ETF structure in meeting liquidity demands for individual investors. To simplify analysis, we assume that the ETF is fully integrated with the underlying stock market, that is, the stock is identical to the ETF and is omitted from the setting. We use the ETF and the underlying stock interchangeably in our discussion depending on the context.[8]

[8]In reality, an ETF is a portfolio of stocks. It may have a tracking error relative to the underlying stock index and may have different liquidity (can be either higher or lower) from its component stocks. The possibility of arbitrage between the ETF and its underlying stocks guarantees a small tracking error and partially alleviates this concern. Moreover, the main objective of this paper is to compare the relative efficiency of OEFs and ETFs rather than to compare their absolute efficiency relative to the underlying stocks. We leave the detailed comparison between the ETF and its underlying stocks to future research.

We model the OEF structure in the following reduced-form way: At time 0, OEF investors give their stock endowments to the fund in exchange for OEF shares. Each OEF share is normalized to be equal to a share of the stock.[9] At time 1, any OEF investor can redeem or purchase any number of shares at a pre-specified price of P_1^F per share, which is independent of the market condition at time 1. The fund can borrow or invest at the risk free rate to accommodate the fund flow. That is, if in aggregate OEF investors demand extra shares at time 1, the fund issues new shares in exchange for cash payments from its investors and then parks the money in the risk free asset; if instead the aggregate demand is a sell order, the fund borrows money and pays out to the departing investors at the price P_1^F. Shortly after that, at time 1_+, the fund passes the demand to the underlying stock market. That is, they buy or sell stock shares at the market price P_1^E to ensure that each remaining OEF share has the same risk exposure as one share of stock. In this sense, the OEF is extremely passive – it does not optimally manage the cash position to account for potential price impacts. To the extent that $P_1^E \neq P_1^F$, the fund makes or loses money on these transactions. The gains or losses are shared evenly by all remaining shares in the OEF, leading to a terminal payoff different from that of the stock. We term the difference the "tracking error" of the OEF.

As will become obvious later, the price difference between P_1^F and P_1^E is predictable at time 1. To rule out arbitrage, we assume that once investors choose to invest in an OEF, they cannot access the ETF or the stock market directly.

3.2 Agents

There is a continuum of investors in the economy with a total population weight of $1+\mu$, of which μ fraction are stock investors and the remaining are index investors choosing between the OEF and the ETF. The focus of the paper is the decision of indexers between the two indexing vehicles. Stock investors are introduced so that there is always a viable stock market in which the OEF can transact (at time 1_+) to implement the demand from its investors, even when the population of ETF investors goes to 0. In addition, it is realistic since indexing demand is less than 20% (which corresponds to a μ of 4) for most stocks.

Each investor is endowed with $\bar{\theta}$ shares of the stock at time 0. Investor i also receives a

[9] In practice, to start investing in an OEF, investors pay cash and the fund purchases stocks. We assume OEFs are formed by accepting stocks directly from their investors to eliminate the impact of OEF investing on time 0 stock prices. This price impact complicates the analysis and does not affect our main message.

non-traded payoff N_i at the terminal date 2, which is given by

$$N_i = Y_i\,(V - \bar{V}), \qquad Y_i = Y + \epsilon_i, \tag{2}$$

where V is the stock payoff in (1), and Y and ϵ_i are mutually independent, normally distributed random variables with a mean of zero and a volatility of σ_Y and σ_i, respectively.[10] We can interpret this non-traded income as a liquidity shock that is correlated with the stock; in particular, it is equivalent to an endowment shock of Y_i shares of the stock. Given the correlation between the liquidity shock and the stock payoff, investors want to adjust their stock positions in order to hedge this risk, giving rise to their trading needs. In particular, stock investors trade in the stock market directly to offset their liquidity shocks while indexers trade either in the OEF or in the ETF depending on their choice of the index vehicle at time 0.[11]

Since all investors are subject to the same Y shock, it captures the aggregate liquidity shock. The component ϵ_i is independent across all investors and defines their individual shock. Some investors face more individual liquidity shocks than others, so we assume

$$\sigma_i = s_i\,\sigma_\epsilon, \quad s_i \sim Unif[0,1], \tag{3}$$

where σ_ϵ is the maximum individual liquidity shock and s_i is the magnitude of liquidity needs for individual i, which is uniformly distributed over the range $[0,1]$. Since ϵ_i is independent with bounded variance and each individual investor has zero population weight, Law of Large Number holds. Thus, for any subset $\mathcal{I}$ of investors, individual liquidity shocks always cancel out and the total liquidity demand depends only on the aggregate shock and is proportional to the population weight $\mu_{\mathcal{I}}$,

$$\int_{i\in\mathcal{I}} \epsilon_i = 0, \qquad \text{and} \qquad \int_{i\in\mathcal{I}} N_i = \mu_{\mathcal{I}}\, Y(V - \bar{V}). \tag{4}$$

All agents have identical preference, which can be described by an expected utility function over the terminal wealth. For tractability, we assume that the agent exhibits mean-

[10] We assume that the non-traded payoff N_i is perfectly correlated with the stock payoff V. As long as N_i and V are correlated, the qualitative nature of our results is independent of the sign and the magnitude of the correlation.

[11] Heterogeneity in endowment is merely a device to introduce the need of trading for risk-sharing purposes as in Diamond and Verrecchia (1981) and Huang and Wang (2008). Other forms of heterogeneity can also generate trading needs, such as difference in preferences or beliefs. Our modelling choice is mainly motivated by tractability.

variance preference. In particular, agent i has the following utility function:

$$\mathrm{E}[W_2^i] - \frac{\gamma}{2}\mathrm{Var}[W_2^i], \tag{5}$$

where γ is the risk aversion and W_2^i is the agent's terminal wealth.

3.3 Time line

We now describe in detail the timing of events and actions. At time 0, indexers decide whether to invest via the ETF or the OEF. The market equilibrium determines the fractions η and $1-\eta$ of investors who choose to invest in the ETF and the OEF, respectively. ETF investors do nothing at time 0, and OEF investors exchange all their stock shares for an equal number of OEF shares.

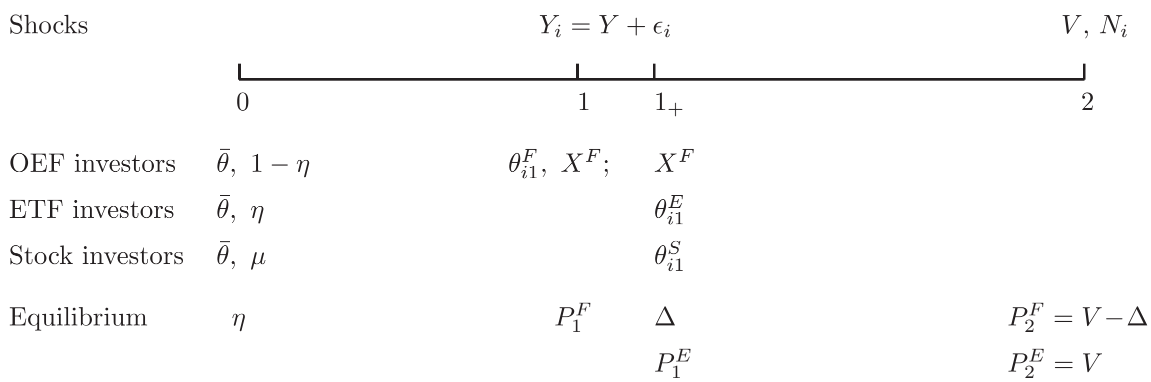


Figure 2: The time line of the economy.

At time 1, investors in the OEF learn the aggregate liquidity shock Y and their individual risk exposure ϵ_i. They take as given the price P_1^F and choose their optimal holding θ_{i1}^F. Investor i redeems $(\bar{\theta} - \theta_{i1}^F)$ shares from the OEF and receives a payment of $P_1^F(\bar{\theta} - \theta_{i1}^F)$ (or pays cash to purchase shares if $(\bar{\theta} - \theta_{i1}^F) < 0$). Aggregating over all OEF investors, the OEF experiences a total redemption of

$$X^F \equiv \int_{i \in OEF} (\bar{\theta} - \theta_{i1}^F) = (1-\eta)\,\bar{\theta} - \int_{i \in OEF} \theta_{i1}^F \tag{6}$$

shares (or creation if $X^F < 0$). Hence, the total OEF shares outstanding is $\int_i \theta_{i1}^F$. The assets of the OEF consist of $(1-\eta)\bar{\theta}$ shares of the stock and $-X^F P_1^F$ dollars of cash.

Note that the supply of OEF shares is not fixed at time 1 and is a function of the predetermined price P_1^F, since the fund can create or redeem any number of shares based on investor demand. In this sense, P_1^F is not an equilibrium price that clears the market, but rather a contractual price that all OEF investors agree upon as a part of the OEF contract. In the model, we impose an additional restriction that the aggregate trading demand from OEF investors (X^F) is zero when the aggregate liquidity shock is $Y = 0$ to pin down the price.

At time 1_+, the OEF passes the aggregate demand from its investors through to the underlying asset market by buying or selling stocks. It sells exactly X^F shares of the stock (independent of the price P_1^E) to make sure that each of the remaining OEF shares has the same risk exposure as a share of the stock. Since the OEF purchases X^F shares from its investors at the price P_1^F and then unloads them in the ETF market at the price P_1^E, it incurs a loss of $X^F(P_1^F - P_1^E)$. This loss is shared equally among all remaining shares, which is $\int_i \theta_{i1}^F$. Hence, the per share loss is

$$\Delta = X^F(P_1^F - P_1^E) / \Big(\int_{i \in OEF} \theta_{i1}^F \Big). \tag{7}$$

At time 1_+, in the ETF market, all stock and ETF investors participate both to unload their own shock Y_i and to accommodate the above demand X^F from the OEF. To simplify notation, we use the same time index 1 for all variables realized at 1_+. In particular, we use P_1^E to denote the equilibrium ETF/stock price and θ_{i1}^E (or θ_{i1}^S) to denote the holdings of an ETF (or stock) investor after trading at time 1_+. There is little room for confusion, since we assume that OEF investors are not allowed to trade in the ETF. This assumption is needed to rule out arbitrage opportunities given the price difference between P_1^F and P_1^E.

At time 2, investors liquidate their OEF or ETF holdings at prices P_2^F and P_2^E, respectively. Obviously, $P_2^E = V$. The OEF has a per share loss of Δ dollars in addition to its stock holding. Hence, $P_2^F = V - \Delta$. We refer to Δ as the tracking error of the OEF relative to the ETF. Including their non-traded labor income in (2), the terminal wealth for OEF and ETF investors are, respectively,

$$W_{i2}^E = \bar{\theta}\, P_1^E + \theta_{i1}^E (V - P_1^E) + Y_i\,(V - \bar{V}) \tag{8a}$$

$$W_{i2}^F = \bar{\theta}\, P_1^F + \theta_{i1}^F (V - \Delta - P_1^F) + Y_i\,(V - \bar{V}). \tag{8b}$$

3.4 Discussions of the model

In this subsection, we provide additional discussions and motivations about several important features of the model.

First, since we equate ETFs with the underlying stocks, all our discussion of the benefit and the cost of the OEF structure should be interpreted as a relative statement in relation to the ETF structure. For example, the diversification benefit of indexing is common to both structures and does not affect the comparison. Also, even though historically ETFs are introduced after OEFs and the natural question to ask is whether ETFs add value, technically our model asks the opposite question of whether OEFs add value above and beyond what ETFs can offer.

Second, the specification of the OEF trading mechanism, while reasonable, is only one of many ways we can specify the OEF contract. For example, in practice, the P_1^F for most mutual funds is the NAV calculated using the daily closing prices of their stock holdings at 4pm. This pricing of the OEF guarantees that, if an investor puts in an order close to 4pm, he will not bear the cost of the future price impact that the OEF may experience when it passes on the aggregate demand (X^F) to the stock market the following day. In the context of the model, this effect translates to a price P_1^F that does not fully incorporate the price impact of the aggregate shock Y. For simplicity, we assume a price P_1^F that is totally independent of Y. Partial dependence of P_1^F on Y does not change the qualitative nature of our result.

Third, we assume there is a unique price P_1^E at which all the ETF investors transact with the aggregate OEF demand (X^F). In practice, for any given shock Y, ETF investors may trade sequentially at different prices depending on their time of entry into the market. The OEF may also transact in the underlying stock market at different prices throughout the day either in anticipation of future demands from its investors or in an effort to implement existing demands. Our model considers a batch market that pools all these demands together to reach a single equilibrium price P_1^E. This provides a clean benchmark to compare the structural difference between the OEF (which provides a partial insurance against the Y risk through the pricing mechanism P_1^F) and the ETF (which provides no insurance). Clearly, our model underestimates the benefit of the ETF structure for informed investors who expect to trade ahead of the aggregate demand. On the other hand, to the extent that no investor has superior information, as we assume in this model, this batch assumption overestimates the benefit of the ETF structure by smoothing the utility of all ETF investors across the different transaction prices they may receive.

Fourth, in reality, ETF investors need to pay bid-ask spreads to the stock exchange whereas the OEF charges zero bid-ask spread by providing internal crossing of individual trades. The identical price P_1^E ignores this difference and overstates the benefits of the ETF, which is again conservative for our purpose. In addition, whenever there is an aggregate liquidity demand, the OEF cannot perfectly cross all its trades and needs to pass the aggregate demand to the underlying stock market. As discussed earlier, the OEF provides partial insurance by offering the traders a price P_1^F that does not fully incorporate the shock and at the same time passing the price discrepancy between P_1^E and P_1^F to the remaining investors. This price discrepancy is the source of costly fund flows identified in the mutual fund literature like Edelen (1999) and Greene and Hodges (2002), among others. When the remaining investors are exceedingly concerned about this negative externality, they may choose to sell shares even when they personally do not experience liquidity shocks and may lead to a "bank run" on the fund as suggested by Chen, Goldstein, and Jiang (2007).

Fifth, we abstract away some institutional details to highlight the structural difference between OEFs and ETFs. For example, an OEF can replicate an index with a small number of stocks while an ETF has to hold the whole index. In addition, during index membership changes, for example, to accommodate the inclusion of Google into the S&P 500 index, an ETF needs to proportionally sell 499 stocks to purchase Google whereas most OEFs can implement this transition more efficiently via fund flow management. This gives the OEF an advantage in transactions costs during index changes. In addition, while an ETF is generally more tax-efficient due to its ability to pass out low-tax-basis stocks to departing investors via the in-kind redemption feature, it may become less tax-efficient when the fund is forced to trade frequently to accommodate index changes.

4 Equilibrium

We solve the equilibrium backwards in three steps. First, taking the demand from the OEF as given, we solve the optimization problem of ETF investors to derive the equilibrium price at time 1_+. Second, we solve the trading decision of OEF investors at time 1 and the equilibrium tracking error. Third, we evaluate the utility of investing via the OEF and the ETF for investors with different liquidity needs at time 0 and determine the equilibrium fraction of investors who choose to invest in each.

4.1 Equilibrium at time 1_+ in the ETF market

Assume the OEF is selling X^F shares of the stock. The following proposition solves the optimal decision of ETF investors and the market equilibrium at time 1_+.

Proposition 1. *Given the population η (and μ) of ETF (and stock) investors, the demand X^F from OEF investors, and the aggregate liquidity shock Y, the equilibrium ETF price at time 1_+ is*

$$P_1^E = \bar{V} - \gamma\sigma_v^2\,\bar{\theta} - \gamma\sigma_v^2\left(Y + \frac{X^F}{\eta + \mu}\right), \tag{9}$$

and the optimal holding of an ETF (and stock) investor i is

$$\theta_{i1}^E = \theta_{i1}^S = \frac{\bar{V} - P_1^E}{\gamma\sigma_v^2} - Y - \epsilon_i \tag{10a}$$

$$= \bar{\theta} + \frac{X^F}{\eta + \mu} - \epsilon_i. \tag{10b}$$

By construction, once the ETF investors decide to invest via the ETF, they are identical to stock investors in the model. From equation (10a), the desired holding of investor i depends on the risk-return tradeoff of the stock (the first term) plus a hedging term against the liquidity shock $Y_i = Y + \epsilon_i$. Since Y is the aggregate shock, it affects everyone's desired holding and hence the equilibrium price in (9). In equilibrium, the price P_1^E fully adjusts for the impact of Y and investors no longer choose to unload Y, as (10b) indicates. In summary, the aggregate risk Y affects only the price but not the trading decisions in equilibrium. All ETF and stock investors equally share the aggregate risk Y and they unload only their idiosyncratic risk exposure, ϵ_i, in the market. The excess demand, X^F, from the OEF investors, is an aggregate risk for the ETF and stock investors (with total population weight $\eta + \mu$). They equally share this risk and the equilibrium price incorporates this risk as well.

4.2 Equilibrium at time 1 in the OEF market

We now solve the OEF equilibrium at time 1 in two steps. First, taking as given the functional forms of both the price P_1^F and the tracking error Δ, OEF investors choose their optimal holding conditional on their liquidity shock. Second, we aggregate the demand from OEF investors and solve for equilibrium prices and the tracking error.

The following proposition derives the equilibrium price P_1^F and individual share holdings based on this restriction.

Proposition 2. *Assume the OEF price is $P_2^F = V - \Delta$ at time 2, where $\Delta = \Delta_0 + \Delta_1 Y + \Delta_2 Y^2$ is the tracking error, and that the aggregate trading demand from OEF investors (X^F) is zero when the aggregate shock is $Y = 0$. Then the equilibrium OEF price at time 1 is*

$$P_1^F = \bar{V} - \gamma\sigma_v^2\,\bar{\theta} - \Delta_0, \tag{11}$$

and the optimal holding of an OEF investor i is

$$\theta_{i1}^F = \frac{\bar{V} - \Delta - P_1^F}{\gamma\sigma_v^2} - Y - \epsilon_i, \tag{12a}$$

$$= \bar{\theta} - \frac{\Delta_1 Y + \Delta_2 Y^2}{\gamma\sigma_v^2} - Y - \epsilon_i. \tag{12b}$$

The holding in (12a) is similar to that of (10a), except that the tracking error Δ lowers the expected future payoff of the OEF and reduces the demand for the stock. Similar to the ETF case, ignoring the functional form of the price, investors would like to hedge both the aggregate shocks Y and the idiosyncratic shock ϵ_i. In contrast to the ETF case, however, the equilibrium price P_1^F does not fully account for the impact of Y. In particular, we model a special case in which P_1^F is independent of Y. As a result, OEF investors optimally unload their aggregate risk exposure Y in equilibrium. This is in direct contrast to the case of ETF investors in Proposition 1, who hedge only the idiosyncratic risk ϵ_i. We term this the "moral hazard" effect: since OEF investors are guaranteed a price P_1^F independent of Y, they have the incentive to unload their aggregate risk exposure Y even though in equilibrium it is more efficient for everyone to share the aggregate risk.

We now connect the OEF and the ETF market through the definitions of OEF demand X^F and the tracking error Δ in (6) and (7), respectively. The definition of the tracking error is highly nonlinear. We use Taylor expansion to expand it around Y. As long as Y is small relative to the total supply of the stock, we can drop higher-order terms. In particular, we drop terms higher than the second order and match coefficients of the Y terms to derive the equilibrium in the following Proposition.

Proposition 3. *When Y is small (relative to $\bar{\theta}$),*

(i) at time 2, the equilibrium OEF price is $P_2^F = V - \Delta$, where Δ is the tracking error and

$$\Delta = \Delta_2 Y^2, \qquad \Delta_2 \equiv (1 + \hat{\eta})\gamma\sigma_v^2/\bar{\theta}, \qquad \hat{\eta} \equiv \frac{1 - \eta}{\eta + \mu} \tag{13}$$

(ii) at time 1, the equilibrium prices of the OEF and the ETF are, respectively,

$$P_1^F = \bar{V} - \gamma\sigma_v^2\,\bar{\theta} \tag{14a}$$

$$P_1^E = \bar{V} - \gamma\sigma_v^2\,\bar{\theta} - \gamma\sigma_v^2(1+\hat{\eta})Y - \hat{\eta}\Delta_2 Y^2 \tag{14b}$$

(iii) at time 1, the equilibrium holdings of the OEF and the ETF investors are, respectively,

$$\theta_{i1}^F = \bar{\theta} - \epsilon_i - Y - (1+\hat{\eta})Y^2/\bar{\theta} \tag{15a}$$

$$\theta_{i1}^E = \bar{\theta} - \epsilon_i + \hat{\eta}Y + \hat{\eta}(1+\hat{\eta})Y^2/\bar{\theta}, \tag{15b}$$

and the aggregate demand from OEF investors is $X^F = (\eta+\mu)(x_1 Y + x_2 Y^2)$, where $x_1 = \hat{\eta}$ and $x_2 = \hat{\eta}(1+\hat{\eta})/\bar{\theta}$.

Note that the population weight of OEF investors is $1-\eta$ and the population weight of ETF and stock investors is $\eta+\mu$, hence, $\hat{\eta} = (1-\eta)/(\eta+\mu)$ in (13) defines the relative population weight of the OEF investors. From (15), OEF investors unload all of their aggregate risk exposure (Y) to the ETF market, while ETF investors accommodate the demand. This extra demand causes an additional impact on the price P_1^E and leads to the OEF tracking error Δ. Although collectively costly, individual OEF investors have little incentive to internalize this cost given their guaranteed price P_1^F. We return to the discussion of the tracking error in more detail later.

4.3 The equilibrium size of the ETF industry at time 0

We first calculate the value functions of the OEF and the ETF investors given the fraction of investors who choose to invest in each; then we solve the equilibrium size of the ETF so that potential investors are indifferent between these two investment vehicles.

The following lemma derives the individual value functions of OEF and ETF investors at time 0. To separate the effect of individual optimization from the equilibrium price impact, we start with the partial equilibrium effect when both the tracking error Δ of the OEF and the order flow X^F faced by ETF investors are taken as exogenously given. Later on, we consider the general equilibrium effect by incorporating the definitions of Δ and X^F from Proposition 3.

Lemma 1. *Assume the tracking error of the OEF is $\Delta = \Delta_2 Y^2$ and the ETF market has an demand shock of $X^F = (\eta+\mu)(x_1 Y + x_2 Y^2)$.*[12] *When σ_y is small (relative to $\bar{\theta}$), for an*

[12] To simplify the expressions, we use the knowledge from Proposition 3 that $\Delta_0 = \Delta_1 = x_0$=0.

investor anticipating an idiosyncratic liquidity shock of the size $\sigma_i = s_i\sigma_\epsilon$, the value function of investing in the OEF and the ETF is, respectively,

$$J_{i0}^F = \bar{\theta}\bar{V} - \frac{1}{2\gamma}\big(1+s_i^2k_\epsilon+k_y\big)k_\theta - (1-s_i^2\,k_\epsilon)\,\bar{\theta}\sigma_y^2\Delta_2 + O(\sigma_y^4), \tag{16a}$$

$$J_{i0}^E = \bar{\theta}\bar{V} - \frac{1}{2\gamma}\big(1+s_i^2k_\epsilon\big)(k_y+k_\theta) + \frac{k_y}{2\gamma}(1-k_\theta)x_1^2 - \frac{k_y}{2\gamma}(2x_1+x_1^2+2x_2\bar{\theta})s_i^2k_\epsilon + O(\sigma_y^4), \tag{16b}$$

where $O(\sigma_y^4)$ are terms of the order σ_y^4 or higher, $k_y \equiv \gamma^2\sigma_v^2\sigma_y^2$, $k_\epsilon \equiv \gamma^2\sigma_v^2\sigma_\epsilon^2$, and $k_\theta \equiv \gamma^2\sigma_v^2\bar{\theta}^2$.

The value function exhibits several interesting features. First, if $\Delta = 0$ and $X^F = 0$, then the gain of investing via the OEF (relative to investing via the ETF, $J_{i0}^F - J_{i0}^E$) increases with s_i, the size of the individual liquidity shock. This reflects the increasing benefit of liquidity insurance provided by the OEF structure.

Second, the tracking error (Δ_2) reduces the utility for OEF investors, especially for those with smaller idiosyncratic liquidity needs (s_i). Although this result appears similar to the common perception that investors with low liquidity needs are subsidizing others with higher liquidity needs, the intuition is different. From (4), all idiosyncratic liquidity needs cancel out at the fund level. Hence, the only liquidity need that is costly to the fund is the aggregate liquidity shock Y. It is important to note that since all investors have equal exposure to the aggregate liquidity shock, they contribute equally to the aggregate trading need and the resulting tracking error of the fund, whether they have high or low idiosyncratic liquidity needs.

Third, the demand from OEF investors have dual impacts on the ETF investors. On the one hand, ETF investors make markets for the OEF investors and earn a profit by doing so. This is reflected in the third term in (16b), $\frac{k_y}{2\gamma}(1-k_\theta)x_1^2$. On the other hand, the extra demand makes price P_1^E more volatile and is costly especially for those with high idiosyncratic liquidity needs, as indicated by the negative term after that, $-\frac{k_y}{2\gamma}(2x_1+x_1^2+2x_2\bar{\theta})s_i^2k_\epsilon$.

In summary, these results suggest that individuals with high idiosyncratic liquidity needs are ill-suited to provide liquidity to others via the ETF structure and they bear a high cost of transacting directly at volatile prices; they benefit more from the insurance feature embedded in the OEF structure and suffer less from the tracking error of OEFs. Hence, there is a natural separation between investors with high and low liquidity needs in their preferred investment vehicle. The following lemma confirms this preference.

Lemma 2. *Let η be the fraction of investors who choose to invest via the ETF. When σ_y is small (relative to $\bar{\theta}$), for an investor anticipating an idiosyncratic liquidity shock of the size*

$\sigma_i = s_i\sigma_\epsilon$, the utility gain of investing via the OEF relative to the ETF is

$$G_i^F(s_i,\eta) \equiv J_{i0}^F - J_{i0}^E = \frac{(1+\hat{\eta})^2}{2\gamma}\big(3s_i^2 k_\epsilon - 1 - \frac{1-\hat{\eta}}{1+\hat{\eta}}k_\theta\big)k_y + O(\sigma_y^4). \tag{17}$$

(i) Investor i invests in the OEF if and only if $G_i^F \geq 0$, otherwise, he invests in the ETF;

(ii) G_i^F increases with individuals' idiosyncratic liquidity shock, that is, $(\partial G_i^F)/(\partial s_i) > 0$.

Lemma 2 suggests that individuals with higher liquidity needs prefer the OEF structure relative to the ETF. One might question the viability of the OEF structure in equilibrium if high-liquidity-need investors are their main clientele. Interestingly, since individual liquidity needs cancel out at the fund level, high-liquidity-need investors do not lead to more volatile fund flows or higher flow-induced trading costs. As a result, the OEF is not dominated by the ETF in equilibrium. The following proposition shows the existence of equilibrium in which the marginal investor is indifferent between the two vehicles and derives the equilibrium size of ETFs.

Proposition 4. *When σ_y is small (relative to $\bar{\theta}$), let η solves the following equation,*

$$\eta = \begin{cases} 0, & \text{if } G_i^F(0,0) > 0; \\ 1, & \text{if } G_i^F(1,1) < 0; \\ G_i^F(\eta,\eta) = 0, & \text{o.w.}, \end{cases} \tag{18}$$

then all indexing investors with $s_i \leq \eta$ invests via the ETF and the rest ($s_i > \eta$) invests via the OEF. The equilibrium population of ETF investors is η.

The proof of the proposition is straightforward. When $G_i^F(0,0) > 0$, for any investors with $s_i > 0$, we have $G_i^F(s_i,0) \geq G_i^F(0,0) > 0$ since G_i^F increases in s_i. Hence, at $\eta = 0$ the OEF is preferred by all investors, confirming that $\eta = 0$ is an equilibrium. Similarly, when $G_i^F(1,1) < 0$, we have $G_i^F(s_i,1) \leq G_i^F(1,1) < 0$ for any s_i and ETF is preferred by everyone, thus, $\eta = 1$. The rest of the proof can be found in the Appendix.

While intuitive, the proposition does not give simple conditions on the underlying parameters for the equilibrium. The following corollary provides a more explicit solution.

Corollary 1. *The equilibrium population of ETF investors can be expressed as follows:*

$$\eta = \begin{cases} 0, & \text{if } \mu < (k_\theta - 1)/(k_\theta + 1); \\ 1, & \text{if } k_\epsilon < (1 + k_\theta)/3; \\ \frac{1}{3(1+\mu)k_\epsilon}\big(k_\theta + \sqrt{{k_\theta}^2 + 3k_\epsilon(1+\mu)^2 + 3k_\epsilon k_\theta(\mu^2-1)}\big), & \text{o.w.} \end{cases} \tag{19}$$

The first case of the corollary indicates that the ETF population drops to 0 for small μ. This is a situation with very few stock investors to make markets. If the equilibrium fraction of ETF is small, the price becomes extremely volatile whenever OEF investors unloads their aggregate demand; the volatile price makes it extremely costly to trade directly in the ETF market, and everyone is better off investing via the OEF to smooth their idiosyncratic shocks. The second case states that for small k_ϵ – which corresponds to small idiosyncratic liquidity shocks – all investors prefer the ETF structure. The reason is that the benefit of liquidity insurance provided by the OEF structure is too low to cover the moral hazard cost of inefficiently unloading the aggregate shocks. We leave the discussion of the general case to the next section.

5 Equilibrium properties and empirical implications

We now examine the properties of the equilibrium. In particular, we discuss the determinants of the OEF tracking error and lay out empirical implications regarding the equilibrium size of the ETF for different underlying indexes.

5.1 The tracking error of the OEF

The tracking error of the OEF is defined in Proposition 3. From (13), $\Delta = \Delta_2 Y^2 > 0$, hence the tracking error always reduces the terminal payoff for the OEF. This is consistent with the finding in the literature that fund flows are costly for the OEFs and can reduce fund performance.

Interestingly, from (13), the coefficient $\Delta_2 \equiv (1+\hat{\eta})\gamma\sigma_v^2/\bar{\theta}$ is strictly positive even when $\hat{\eta}$, the relative population of OEF investors, is zero. In this case, given their tiny population weight, OEF investors do not affect the demand of the ETF investors in (15b) or the equilibrium price P_1^E in (14b). Yet there is still a price discrepancy between P_1^E (which depends on Y) and P_1^F (which is independent of Y). In particular, if $Y > 0$, the aggregate liquid shock is equivalent to an excess supply, which reduces P_1^E to below P_1^F. At the same time, the OEF investors unload all their Y risk by selling Y shares of their stocks. Since they are able to sell their shares to the OEF at a price (P_1^F) higher than the OEF can realize in the stock market later (P_1^E), the OEF is losing money on average and the tracking error is negative. Similarly, if $Y < 0$, OEF investors buy shares from the OEF at a price lower than the OEF can purchase in the stock market later, also leading to a negative tracking error.

When $\hat{\eta} > 0$, OEF investors follow similar strategy and unload all their Y risk. Given their non-trivial population size, they now affect the equilibrium demand of ETF investors and the price P_1^E. In particular, each ETF investor purchases $\hat{\eta}$ shares for each share sold by an OEF investor. This extra supply ($\hat{\eta}Y$ on a per capita basis), further depresses the ETF price P_1^E (by an extra $\gamma\sigma_v^2\hat{\eta}Y$). With P_0^F staying constant, the OEF now loses an average of $\gamma\sigma_v^2(1+\hat{\eta})Y$ per share on all the Y shares sold, leading to the tracking error $\Delta_2 Y^2$. This tracking error term further reduces the demand for OEF shares and leads to additional price impact in the ETF market, contributing to the last terms in the expressions of (14b) and (15). Note that for small μ, $\hat{\eta}$ can be very large as η approaches 0. This represents the case when the market is dominated by OEF investors. With very few ETF and stock investors to make markets, the price P_1^E can be extremely sensitive to the aggregate liquidity shock. The OEF needs to pay a significant cost to unload all their Y risk to the small underlying stock market, leading to significant price impacts and tracking errors.

In summary, the tracking error consists of two components. The first component is due to the price discrepancy between the OEF price and the price in the ETF market when we take the ETF price as exogenous. This is the moral hazard cost of providing the liquidity insurance. The second component of the tracking error is a "feedback effect," driven by the price impact in the underlying stock market when OEF investors unload their aggregate risk exposure. The price impact then leads to additional tracking errors for the OEF, further decreasing investors' demand for the OEF.[13] This component of tracking error increases with the size of the OEF. Thus, we have the following result.

Result 1. *The tracking error always reduces the OEF performance, and it is not zero even when the size of the OEF approaches zero. Moreover, the magnitude of the tracking error increases as the relative size of the OEF ($\hat{\eta}$) increases.*

5.2 The cross-section of the OEF demand

Our framework allows us to assess the effectiveness of the OEF and the ETF structures under different circumstances. Viewing the underlying risky asset in the model as a stock index, we can compare characteristics of different stock indexes and ask whether one index

[13] In the paper we only consider the case of small Y for tractability. In practice, when Y is reasonably large, this effect can manifest into a "bank run" situation, as considered in Chen, Goldstein, and Jiang (2007).

is more suited for the OEF or the ETF investors. In Figure 3, we report the equilibrium size of the ETF relative to the OEF for different indexes and investor characteristics.

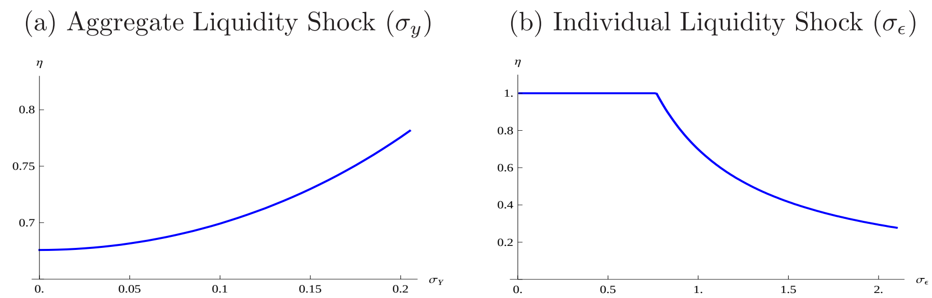


Figure 3: **The equilibrium size of the ETF industry (η).** Panels (a) and (b) report the optimal size of ETF for different values of σ_y and σ_ϵ, respectively. Other than the variables being changed, the parameters are $\gamma = 4$, $\bar{\theta} = 1$, $\bar{V} = 0$, $\bar{Y} = 0$, $\sigma_v = 0.3$, $\sigma_y = 0.1$, $\sigma_\epsilon = 1.0$, and $\mu = 1$.

First, we consider the impact of aggregate liquidity shock. A more volatile aggregate liquidity shock (higher σ_y) corresponds to the case in which investors have more correlated liquidity needs. We have the following result:

Result 2. *When the aggregate liquidity shock is more volatile, it is more costly to provide liquidity insurance via the OEF structure, and the equilibrium size of the ETF is larger.*

With more volatile aggregate liquidity shock, the OEF is expected to have larger unbalanced demand from its investors. This unbalanced demand is going to translate into large excess demand in the underlying stock market, leading to a large price impact and a large tracking error. As a result, the OEF becomes less attractive relative to the ETF as an indexing vehicle, and the equilibrium ETF size is larger.

Empirically, it is reasonable to assume that investors of a narrower index (such as a biotech index) have more correlated liquidity needs compared to investors of a broader market index (such as the S&P 500 index). Thus, we expect to have more ETFs in narrower indexes than in broader market indexes. We proxy for the narrowness of an index using the number of stocks, the market capitalization, and the industry concentration of the index.

Next, we consider the impact of idiosyncratic liquidity shocks. A more volatile idiosyncratic liquidity shock (higher σ_ϵ) corresponds to the case in which investors on average have large liquidity needs but these liquidity needs are not correlated. We have the following result:

Result 3. *When the individual liquidity shock is more volatile, the liquidity insurance provided by the OEF structure is more beneficial, and the equilibrium size of the ETF is smaller.*

Figure 3(b) indicates that if all investors in the economy face relatively low idiosyncratic shocks (low σ_ϵ), the equilibrium size of ETF can reach 1, or there is no need for the OEF. The reason is that investors with low idiosyncratic liquidity demand can still generate significant tracking errors as long as they are exposed to the aggregate liquidity shock. On the other hand, the liquidity insurance feature – the key benefit of the OEF structure – is not very important for them. As a result, the ETF structure dominates as an indexing vehicle.

This result points out a potential pitfall of the recent trend in the mutual fund industry to impose trading restrictions to deter high frequency trading in the OEFs. Our findings suggest that this strategy may not be efficient since high frequency traders do not pose high costs to the fund, unless there is reason to believe that their trading is more correlated with the aggregate shocks. On the other hand, those investors benefit much more from the liquidity insurance provided by the OEF structure and hence are the natural demanders of the OEFs. If they are deterred from investing in the OEFs by the trading restrictions, then OEFs can potentially lose a significant client base.

Empirically, we expect investors to use more liquid assets to meet their individual liquidity needs, hence, those indexes are likely to have a client base with larger idiosyncratic liquidity needs. Thus, Result 3 suggests that less liquid underlying indexes should have a client base with smaller idiosyncratic liquidity needs and the equilibrium size of the ETF is larger.

In addition, we can proxy for the idiosyncratic liquidity needs (σ_ϵ) using the characteristics of fund investors. For example, investors of retirement accounts or other institutional accounts generally have lower liquidity needs than the average individual investors. Our model suggests they may be better candidates for ETFs since they do not highly value the liquidity insurance of OEFs.

In summary, we find that narrower indexes (e.g., indexes with fewer stocks or are more concentrated) and less liquid indexes are better suited for ETFs. Hence, we predict that these are the indexes in which more ETFs are introduced and, once introduced, ETFs enjoy a higher growth rate. In contrast, the corresponding OEFs should experience a bigger loss of market share to ETFs once these ETF are introduced.

6 Data and empirical analysis

6.1 Data description

We use several different data sets in generating our sample:

First, we use CRSP's *Survivor-Bias Free U.S. Mutual Funds Database*, from which we extract a list of the index mutual funds (OEFs). We define a fund to be an index mutual fund if the word index (or any derivative of it such as Ind, Idx, and so on) shows up in its name. Whenever in doubt, we check the prospectus of the fund using *EDGAR*, the *SEC Filings and Forms Database*, in order to verify that indeed it is a pure index fund. This step of verification from EDGAR is necessary, as some mutual funds can track an index closely (and have the word index in their name) and yet not be an index fund. For example, self described "enhanced index" funds essentially track an index but knowingly change some weights of stocks they believe will over- or under-perform the index, and hence "enhance" the index. If the prospectus is not clear that the fund performs only the task of tracking an index, we err on the side of caution and drop the fund from our sample. We then merge the different share classes of each fund into one entity. As has been recognized in the literature, the same mutual fund can have several almost identical share classes. The different share classes usually differ either in their fee structure (i.e., front-end load, back-end load, no load, and so on) or in their distribution channel (i.e., investor class, institutional class, and so on). We merge the different share classes into one OEF entity by value weighting the different classes based on their Total Net Asset (TNA). The resulting sample is comprised of 296 OEFs over the time span of 1992-2006. These funds are managed by 131 different families. Interestingly, these 296 funds track only 63 different indexes.

Second, we use Bloomberg in order to generate a list of ETFs. We collect the ticker and CUSIP of each ETF, we also find out what fund family is sponsoring it. In order to verify that we have captured the full span of ETFs, we check the websites of the respective sponsoring families and ETF-dedicated websites (such as ETFConnect.com). The resulting sample is comprised of 320 ETFs over the time span of 1992-2006. We restrict both our ETF and OEF samples to start in 1992 due to the fact that there were no ETFs before then. This allows us to capture the entire current growth of ETFs from their creation to 2006. We exclude any ETF that existed less than one year, as we would not have enough information about it. These 320 ETFs are sponsored by 23 separate families, and they track 268 different indexes. Thus, the final sample consists of 296 OEFs and 320 ETFs over 1992-2006, tracking

289 different indexes.

Third, we collect information about the ETFs and OEFs such as assets under management, number of shares outstanding, creation and deleting date, from CRSP and CRSP Mutual Fund merging by ticker and year.

Last, in order to analyze the indexes tracked by these funds, we need their historical compositions. These compositions are not available for all the indexes. Hence, we employed a proxy for their composition. We use *Thomson Financials*'s *CDA/Spectrum Mutual Funds Holding* and *CDA/Spectrum Institutional (13f) Holdings*. For each index we find all the OEFs and ETFs that track it, since they have to report their holdings on a quarterly basis. Based on the holdings we calculate the variables relating to volatility and composition of the index.

6.2 Summary statistics

Tables 1 and 2 give the summary statistics of the OEFs and ETFs that comprise the dataset. Table 1 reports the summary statistics of the entire dataset. However, since the universe of ETFs and OEFs has been quickly changing, we describe the time trends in the data in Table 2. Though introduced in the 1970s, by the 1990s index OEFs were still not very popular. In 1991 there were only 34 OEFs offered by 19 families managing about 11 billion dollars. ETFs entered the market in 1993 when there was a large inflow in index OEFs. As mentioned in Section 2, though introduced in 1993, ETFs took several years to get exposure. In 1996 assets under management in OEFs had increased to more than 91 billion dollars and ETFs had 3 billion under management.

By 2001, the trend of growth had flipped between OEFs and ETFs. In 2001 ETFs were already growing at a rate three times faster than OEFs. The average growth rate in 2001 of index OEFs was only 7% and of ETFs was 389%. Several interesting trends can be observed in the data. Both the number of families offering OEFs and offering ETFs increased between 1996 and 2006, from 40 to 72 for the OEFs and from 2 to 23 for the ETFs. This highlights two interesting facts. First, the rise in the interest in indexing in the late 1990s brought an increase in offering of index funds by many families. Second, relatively few families offered ETFs even though they were growing at a much faster pace than OEFs, and even though offering one or the other is not technically very different. Vanguard is an example of a family that essentially offers mainly index OEFs and yet refused for many years (under the influence of John Bogle) to also offer ETFs. These facts explain the much higher growth in assets under management per family in ETFs than in OEFs. As we mentioned earlier, OEFs are not always available to all investors depending on the family's distribution channel, making it more difficult at times to grow. By being available to all investors with a brokerage account, ETFs have enjoyed an advantage in growth that is fairly apparent in the data.

Another interesting fact is the number of indexes tracked offered by each sponsoring family. This number has steadily been increasing for ETF sponsors, from 9 per family in 1996 to more than 13 in 2001. During the same time period OEF sponsors have only marginally increased the number of different indexes covered from 2 in 1996 to 2.3 in 2001. This highlights that there might be two different forms of growth not shared by the two organizational forms that is captured by the number of covered indexes offered.

6.3 Indexes tracked

As we show in Tables 1 and 2 a large part of the growth of the ETF industry is through the offering of many different ETFs by each family, each one tracking a different index. In Section 3, we argue that ETFs are better suited than OEFs to handle less liquid and less diversified indexes. Hence, we expect that this growth will not only be by offering many more indexes but also be concentrated in less diversified and less liquid indexes. In order to investigate this implication, we calculate characteristics of these indexes in order to assess this implication. In Table 3 we describe the characteristics of the different indexes. The diversity across different indexes is quite large, they vary from tracking more than two thousand stocks to as few as five stocks. The market capitalization of these indexes is also quite diverse, ranging from small "niche" indexes tracking a segment of 35 billion dollars to market-wide indexes tracking more than 1 trillion dollars. We look at four main characteristics of these indexes: the size of the index, the number of index changes, the liquidity of the underlying stocks, and the concentration of the industry tracked. As expected, indexes vary significantly on all these dimensions. One of the larger costs associated with tracking an index is the cost of re-balancing. The average number of stocks entering/exiting the index is 18 stocks a quarter, with an average market value of 119 billion dollars, so these costs can be very high. The tracking error generated by re-balancing the index is higher for less liquid and smaller stocks. There is also significant variation in the required re-balancing across indexes – some indexes require almost no re-balancing while others require selling and buying more than 60 stocks in a quarter.

We use Amihud (2002)'s liquidity measure. We calculate it by downloading the measure for all stocks from Joel Hasbrouck's website (which is described in detail in Hasbrouck (2006)) and then averaging the underlying liquidity measures of the stocks comprising the index, both as a value-weighted average and as an equal-weighted average. Again, the difference across indexes is fairly substantial. Some indexes have relatively high liquidity stocks (the S&P 500, for example) and some have fairly low liquidity (some industry specific indexes, for example). Since the liquidity of the underlying index directly affects the tracking error of the OEF, we know from our theoretical argument that the best vehicle (OEFs or ETFs) tracking a particular index is dependent on the index liquidity. We calculate the level of industry concentration using two different measures: we use a Herfindahl-Hirschman Index and the industry concentration measure developed in Kacperczyk, Sialm, and Zheng (2007). As we can see in Table 3, there is variation on this level too.

Table 3 shows that indeed as implied by our model the indexes tracked by ETFs and OEFs are different on average. ETFs track smaller, narrower, more concentrated, and less liquid indexes.

7 Conclusions

Fifteen years ago, investors seeking to invest their money by tracking an index could only do so by buying an OEF. In the past decade this paradigm has changed with the introduction of ETFs that are similar to OEFs but are traded on the stock exchange. The huge success of ETFs, seemingly at the expense of OEFs, raises a very interesting question regarding the advantage and disadvantage of the organizational form of an OEF from the investor's point of view.

To address this question, we develop a theoretical model of equilibrium choice between these two vehicles. We show that OEFs provide cross-subsidization among investors by sharing the transaction costs for those investors experiencing large liquidity shocks, at the cost of lower average returns for the remaining investors. While it is a zero-sum game, the OEF structure provides partial insurance against future liquidity needs and is ex-ante beneficial for risk averse investors. However, the insurance feature embedded in the OEF structure can cause moral hazard issues in the form of excessive trading and increase the cost of the insurance. We also find that investors with higher liquidity needs benefit more from the liquidity insurance and hence prefer to invest via the OEF. Interestingly, the concentration of high-liquidity-need investors in the OEF does not lead to higher flow-induced trading costs since individual liquidity needs cancel out at the fund level. As a result, OEFs are not dominated by ETFs in equilibrium. Moreover, we predict that the ETF is a more suitable investment vehicle when investors have more correlated liquidity shocks or when the underlying indexes are narrower or less liquid.

Based on the implications of our theoretical model, we explore the existing data. We find that although ETFs are growing at a rate twice as high as the growth rate of OEFs, part of this growth is explained by a much larger and wider offering of indexes. ETFs cover five times as many indexes as OEFs. OEFs tend to cover the larger, more diversified, more liquid, and more representative indexes. ETFs, in addition to covering those indexes, also offer many other less liquid, less diversified, and more volatile indexes, which our theoretical model shows would be very costly in terms of a "tracking error" to offer as an OEF.

This paper is the first seeking to understand the growth of the ETF industry. Our main finding is that although this industry is growing and will probably continue to grow in the future, it is not likely to totally replace the standard OEFs. ETFs are a new and exciting addition to the different investment options available both to individual and institutional investors. However, our model demonstrates that one cannot view all ETFs as equal. Though some ETFs are very similar to OEFs and offer investors a fairly identical investment vehicle (except for the liquidity cross-subsidization), other ETFs offer new options to track narrower and less liquid indexes that are not cost effective in the OEF structure.

References

Agapova, A., 2006, "Innovations in Financial Products. Conventional Mutual Funds versus Exchange Traded Funds," Florida Atlantic University Working Paper.

Amihud, Y., 2002, "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects," *The Journal Financial Markets*, 5, 31–56.

Berk, J. B., and R. C. Green, 2004, "Mutual Fund Flows and Performance in Rational Markets," *Journal of Political Economy*, 112, 1269–1295.

Berk, J. B., and R. Stanton, 2007, "Managerial Ability, Compensation, and the Closed-End Fund Discount," *The Journal of Finance*, 62, 529–556.

Chalmers, J. M. R., R. M. Edelen, and G. B. Kadlec, 2001, "On the Perils of Intermediaries Setting Security Prices: The Mutual Fund Wildcard Option," *Journal of Finance*, 61, 2209–2236.

Chay, J., and C. A. Trzcinka, 1999, "Managerial Performance and the Cross-Sectional Pricing of Closed-End Funds," *Journal of Financial Economics*, 52, 379–408.

Chen, J., H. Hong, M. Huang, and J. D. Kubik, 2004, "Does Fund Size Erode Mutual Fund Performance? The Role of Liquidity and Organization," *American Economic Review*, pp. 1276–1302.

Chen, Q., I. Goldstein, and W. Jiang, 2007, "Payoff Complementarities and Financial Fragility: Evidence from Mutual Fund Outflows," Working Paper.

Cherkes, M., J. Sagi, and R. Stanton, 2007, "A Liquidity-Based Model of Closed-End Funds," *The Review of Financial Studies*, Forthcoming.

Chordia, T., 1996, "The Structure of Mutual Fund Charges," *Journal of Financial Economics*, 41, 3–39.

Christoffersen, S. E. K., D. B. Keim, and D. K. Musto, 2007, "Valuable Information and Costly Liquidity: Evidence from Individual Mutual Fund Trades," McGill University and University of Pennsylvania Working Paper.

Christoffersen, S. E. K., and D. Musto, 2002, "Demand Curves and the Pricing of Money Mangement," *The Review of Financial Studies*, 15, 1499–1524.

Diamond, D. W., and P. H. Dybvig, 1983, "Bank Runs, Deposit Insurance and Liquidity," *Journal of Political Economy*, 91, 401–419.

Diamond, D. W., and R. E. Verrecchia, 1981, "Information Aggregation in a Noisy Rational Expectations Economy," *Journal of Financial Economics*, 9, 221–235.

Dickson, J. M., J. B. Shoven, and C. Sialm, 2000, "Tax Externalities of Equity Mutual Funds," *National Tax Journal*, 53, 607–628.

Edelen, R. M., 1999, "Investor Flows and the Assessed Performance of Open-end Fund Managers," *Journal of Financial Economics*, 53, 439–466.

Edelen, R. M., R. Evans, and G. B. Kadlec, 2007, "Scale effects in mutual fund performance: The role of trading costs," Boston College, University of Virginia, and Virginia Tech Working Paper.

Elton, E. J., M. J. Gruber, G. Comer, and K. Li, 2002, "Spiders: Where Are the Bugs?," *The Journal of Business*, 75, 453–472.

Gastineau, G. L., 2004, "Protecting Fund Shareholders From Costly Share Trading," *Financial Analyst Journal*, May/June, 22–32.

Goetzmann, W. N., Z. Ivkovic, and G. K. Rouwenhorst, 2001, "Day Trading International Mutual Funds: Evidence and Policy Solutions," *The Journal of Financial and Quantitative Analysis*, 36, 287–309.

Greene, J. T., and C. W. Hodges, 2002, "The Dilution Impact of Daily Fund Flows on Openend Mutual Funds," *Journal of Financial Economics*, 65, 131159.

Grinblatt, M., and S. Titman, 1989, "Mutual Fund Performance: An Analysis of Quarterly Portfolio Holdings," *The Journal of Business*, 62, 393–416.

Hasbrouck, J., 2006, "Trading Costs and Returns for US Equities: Estimating Effective Costs from Daily Data," New York University Working Paper.

Huang, J., and J. Wang, 2008, "Market Liquidity, Asset Prices, and Welfare," *Journal of Financial Economics*, forthcoming.

Johnson, W. T., 2004, "Predictable Investment Horizons and Wealth Transfers among Mutual Fund Shareholders," *Journal of Finance*, 59, 19792011.

Kacperczyk, M., C. Sialm, and L. Zheng, 2007, "Unobserved Actions of Equity Mutual Funds," *The Review of Financial Studies*, Forthcoming.

Lee, C. M. C., A. Shleifer, and R. H. Thaler, 1991, "Investor Sentiment and the Closed-End Fund Puzzle," *The Journal of Finance*, 46, 75–109.

Massa, M., 1997, "Why So Many Mutual Funds? Mutual Funds, Market Segmentation and Financial Performance," INSEAD and CEPR Working Paper.

Nanda, V., M. P. Narayanan, and V. A. Warther, 2000, "Liquidity, Investment Ability, and Mutual Fund Structure," *Journal of Financial Economics*, 57, 417–443.

Pontiff, J., 1996, "Costly Arbitrage: Evidence from Closed-End Funds," *The Quarterly Journal of Economics*, 111, 1135–1151.

Poterba, J. M., and J. B. Shoven, 2002, "Exchange-Traded Funds: A New Investment Option for Taxable Investors," *American Economic Review*, 92, 422–427.

Shleifer, A., and R. W. Vishny, 1997, "The Limits of Arbitrage," *The Journal of Finance*, 52, 35–55.

Stein, J. C., 2005, "Why are most Funds Open-end? Competition and the Limits of Arbitrage," *The Quarterly Journal of Economics*, 120, 247–272.

Viceira, L. M., and A. B. Wagonfeld, 2007, "Barclays Global Investors and Exchange Traded Funds," *HBS Cases*, N9-208-033.

Zitzewitz, E., 2006, "How Widespread was Late Trading in Mutual Funds?," *American Economic Review*, 96, 284–289.

A Appendix

Proof of Proposition 1

Plugging $P_2^E = V$ into the definition of terminal wealth W_{i2}^E for an ETF investor in (8a) and integrating over the distribution of V at given Y_i and θ_{i1}^E, we have

$$\mathrm{E}[W_{i2}^E \mid Y_i] = \bar{\theta} P_1^E + \theta_{i1}^E(\bar{V} - P_1^E) \tag{A1a}$$

$$\mathrm{Var}[W_{i2}^E \mid Y_i] = (\theta_{i1}^E + Y_i)^2 \sigma_v^2 \tag{A1b}$$

Thus, the expected utility J_1^E at time 1_+, defined as the utility (5) conditional on Y_i is:

$$J_1^E \equiv \max_{\theta_{i1}^E} \left[\bar{\theta} P_1^E + \theta_{i1}^E(\bar{V} - P_1^E) - \frac{1}{2}\gamma\sigma_v^2 (Y_i + \theta_{i1}^E)^2 \right]. \tag{A2}$$

The optimal holding is calculated by solving the first-order condition with respect to θ_{i1}^E,

$$\theta_{i1}^E = \frac{\bar{V} - P_1^E}{\gamma\sigma_v^2} - Y_i. \tag{A3}$$

The holdings for stock investors are identical to those for ETF investors. Plugging $Y_i = Y + \epsilon_i$ and (A3) into the market clearing condition, from (4), we have

$$\int \theta_{i1}^E + \int \theta_{i1}^S = (\eta + \mu) \left(\frac{\bar{V} - P_1^E}{\gamma\sigma_v^2} - Y \right) = (\eta + \mu)\bar{\theta} + X^F. \tag{A4}$$

Solving (A4) yields the equilibrium price P_1^E in (9). The optimal holding in the proposition is obtained by substituting the equilibrium price P_1^E back into (A3).

Proof of Proposition 2

The proof of (12a) is almost identical to that of Proposition 1, except that $\mathrm{E}[P_2^F] = \bar{V} - \Delta$. We then plug the definition of Δ into individual holding (12a) and aggregate over all OEF investors:

$$\int_i \theta_{i1}^F = \int_i \left(\frac{\bar{V} - \Delta - P_1^F}{\gamma\sigma_v^2} - Y - \epsilon_i \right) = (1 - \eta) \left(\frac{\bar{V} - \Delta_0 - \Delta_1 Y - \Delta_2 Y^2 - P_1^F}{\gamma\sigma_v^2} - Y \right).$$

Imposing the restriction that $\int_i \theta_{i1}^F(Y = 0) = (1 - \eta)\bar{\theta}$ yields the expression for P_1^F in the proposition. Substituting this P_1^F back into (12a) to derive the expression in (12b).

Proof of Proposition 3

To calculate the tracking error, we plug the expressions of P_1^E, P_1^F and θ_{i1}^F from Propositions 1 and 2 into the definitions of X^F and Δ in (6) and (7). Since $\int_i \epsilon_i = 0$, we have

$$\Delta = \frac{((1-\eta)\bar{\theta} - \int_i \theta_{i1}^F)(P_1^F - P_1^E)}{\int_i \theta_{i1}^F} \tag{A5a}$$

$$= \frac{Y(\gamma\sigma_v^2 + \Delta_1 + \Delta_2 Y)(-\Delta_0 + ((1+\hat{\eta})\gamma\sigma_v^2 + \hat{\eta}(\Delta_1 + \Delta_2 Y))Y)}{\gamma\sigma_v^2(\bar{\theta} - Y) - (\Delta_1 + \Delta_2 Y)Y} \tag{A5b}$$

When Y is small, we can use Taylor expansion to expand around Y and drop higher order terms (i.e., $O(Y^3)$). Equalizing the coefficients of zero, first, and second order terms in (A5b) to those in the definition of $\Delta \equiv \Delta_0 + \Delta_1 Y + \Delta_2 Y^2$, we get three equations and three unknowns (Δ_0, Δ_1, and Δ_2). The solution is $\Delta_0 = \Delta_1 = 0$ and Δ_2 is given in (13).

Parts (ii) and (iii) are straightforward. Substituting the expressions of Δ_0, Δ_1, and Δ_2 into (11) and (12) yields P_1^F and θ_{i1}^F. Using the definitions of Δ, X^F in (6) and Proposition 1, we derive the equilibrium expressions of P_1^E and θ_{i1}^E.

Proof of Lemma 1

First, we calculate the time 0 expectation and variance of W_{i2}^F conditional on the realization of Y.

$$\begin{aligned}
\mathrm{E}[W_{i2}^F \mid Y] &= \mathrm{E}[\mathrm{E}[W_{i2}^F \mid \epsilon_i, Y]] = \bar{\theta}\bar{V} - Y(\gamma\sigma_v^2 + 2\Delta_2 Y)\bar{\theta} + O(Y^3) \\
\mathrm{Var}[W_{i2}^F \mid Y] &= \mathrm{E}[\mathrm{Var}[W_{i2}^F \mid \epsilon_i, Y]] + \mathrm{Var}[\mathrm{E}[W_{i2}^F \mid \epsilon_i, Y]] \\
&= (1 + s_i^2 k_\epsilon)\bar{\theta}(\sigma_v^2\bar{\theta} - 2\Delta_2 Y^2/\gamma) + O(Y^3)
\end{aligned}$$

We then calculate the unconditional expectation and variance:

$$\begin{aligned}
\mathrm{E}[W_{i2}^F] &= \mathrm{E}[\mathrm{E}[W_{i2}^F \mid Y]] = \bar{\theta}\bar{V} - 2\bar{\theta}\Delta_2\sigma_y^2 + O(\sigma_y^4) \\
\mathrm{Var}[W_{i2}^F] &= \mathrm{E}[\mathrm{Var}[W_{i2}^F \mid Y]] + \mathrm{Var}[\mathrm{E}[W_{i2}^F \mid Y]] \\
&= (1 + s_i^2 k_\epsilon + k_y)k_\theta/\gamma^2 - 2(1 + s_i^2 k_\epsilon)\bar{\theta}\Delta_2\sigma_y^2/\gamma + O(\sigma_y^4)
\end{aligned}$$

Then, $J_{i0}^F = \mathrm{E}[W_{i2}^F] - (\gamma/2)\mathrm{Var}[W_{i2}^F]$ yields (16a).

Similarly, we calculate the time 0 expectation and variance of W_{i2}^E conditional on the realization of Y.

$$
\begin{aligned}
\mathrm{E}[W_{i2}^E \mid Y] &= \mathrm{E}[\mathrm{E}[W_{i2}^E \mid \epsilon_i, Y]] = \bar{\theta}\bar{V} + \gamma\sigma_v^2(\bar{\theta}x_1 Y + (x_1 + x_1^2 + \bar{\theta}x_2)Y^2) + O(Y^3) \\
\mathrm{Var}[W_{i2}^E \mid Y] &= \mathrm{E}[\mathrm{Var}[W_{i2}^E \mid \epsilon_i, Y]] + \mathrm{Var}[\mathrm{E}[W_{i2}^E \mid \epsilon_i, Y]] \\
&= (1 + s_i^2 k_\epsilon)\sigma_v^2(\bar{\theta}^2 + 2\bar{\theta}(1 + x_1)Y + ((1 + x_1)^2 + 2\bar{\theta}x_2)Y^2) + O(Y^3)
\end{aligned}
$$

We then calculate the unconditional expectation and variance:

$$
\begin{aligned}
\mathrm{E}[W_{i2}^E] &= \mathrm{E}[\mathrm{E}[W_{i2}^E \mid Y]] = \bar{\theta}\bar{V} + k_y(x_1 + x_1^2 + x_2\bar{\theta})/\gamma + O(\sigma_y^4) \\
\mathrm{Var}[W_{i2}^E] &= \mathrm{E}[\mathrm{Var}[W_{i2}^E \mid Y]] + \mathrm{Var}[\mathrm{E}[W_{i2}^E \mid Y]] \\
&= (x_1^2 k_y k_\theta + (1 + s_i^2 k_\epsilon)(k_\theta + k_y((1 + x_1)^2 + 2x_2\bar{\theta})))/\gamma^2 + O(\sigma_y^4)
\end{aligned}
$$

Then, $J_{i0}^E = \mathrm{E}[W_{i2}^E] - (\gamma/2)\mathrm{Var}[W_{i2}^E]$ yields (16b).

Proof of Lemma 2

By substituting the definitions of Δ_2, x_1, and x_2 from Proposition 3 into the value functions J_{i0}^F and J_{i0}^E in Lemma 1, we derive

$$
\begin{aligned}
J_{i0}^F &= \bar{\theta}\bar{V} - \frac{k_\theta}{2\gamma}\big(1 + s_i^2 k_\epsilon + k_y\big) - \frac{1 + \hat{\eta}}{\gamma}(1 - s_i^2\, k_\epsilon)\, k_y + O(\sigma_y^4) \\
J_{i0}^E &= \bar{\theta}\bar{V} - \frac{k_\theta}{2\gamma}\big(1 + s_i^2 k_\epsilon + k_y\big) - \frac{(1 + \hat{\eta})(1 + 3\hat{\eta})}{2\gamma} s_i^2\, k_\epsilon\, k_y - \frac{1 - \hat{\eta}^2}{2\gamma}(1 - k_\theta)\, k_y + O(\sigma_y^4).
\end{aligned}
$$

Taking the difference between J_{i0}^F and J_{i0}^E yields (17). Clearly, an individual investor should invest in OEF if and only if $G_i^F \geq 0$.

To prove part (ii), we take derivative of G_i^F:

$$
\frac{\partial G_i^F}{\partial s_i} = \frac{(1 + \hat{\eta})^2}{\gamma} 3 s_i k_\epsilon > 0.
$$

Proof of Proposition 4 and Corollary 1

We have proved the cases of $G_i^F(0, 0) > 0$ and $G_i^F(1, 1) < 0$ of Proposition 4 in the text. Using (17), we can verify that the first two conditions in (19) correspond exactly to those two conditions.

We now consider the case of $G_i^F(0, 0) \leq 0 \leq G_i^F(1, 1)$. Whenever $k_\epsilon \geq (1 + k_\theta)/3$, we can verify that the expression of η in the third case of (19) is bounded between $(0, 1)$ and solves $G_i^F(\eta, \eta) = 0$.

For any $s_i > \eta$, we have $G_i^F(s_i, \eta) \geq G_i^F(\eta, \eta) = 0$, hence s_i invests via the OEF. For any $s_i < \eta$, we have $G_i^F(s_i, \eta) \leq G_i^F(\eta, \eta) = 0$, hence s_i invests via the ETF. Given that $s_i \in Unif[0, 1]$, the population of investors with $s_i < \eta$ is exactly η. Hence, the population of ETF investors is η.

Table 1: Summary Statistics

This table provides descriptive statistics of the full sample of OEFs and ETFs for the years 1992 to 2006. The number are calculated using the full sample including mutual funds that do not exit for the entirety of the time period covered. OEFs are Open-Ended Index Mutual Funds, and all the OEF share classes are collapsed to one entity. ETFs are Exchange Traded Funds tracking indexes and trading on the secondary market. All sizes are in millions of dollars. The percentage growth numbers are winsorized at the 2% level.

Variable		OEF	ETF	All
Number of Funds		296	320	616
Number of Families		131	23	151
Number of Indexes Tracked		63	268	289
Funds per Family	Mean	2.31	11.17	3.36
	Median	1	2	1
	Std	2.67	21.42	8.45
Number of Indexes per Family	Mean	2.34	13.70	4.02
	Median	1	6	2
	Std	3.08	23.75	10.44
Fund Size ($M)	Mean	1,828.55	986.22	1,507.46
	Median	240.40	135.46	193.24
	Std	7,906.04	3,961.83	6,694.74
Family Size ($M)	Mean	4,240.15	11,219.33	5,097.05
	Median	440.47	1,401.91	508.81
	Std	26,737.62	31,004.52	28,099.72
Percentage Family Growth	Mean	36.89	113.39	43.14
	Median	15.75	67.25	19.09
	Std	83.06	182.02	90.07
Percentage Fund Growth	Mean	41.32	121.45	69.32
	Median	21.73	55.79	29.99
	Std	79.36	204.34	134.41

Table 2: Summary Statistics - Time Trend

This table provides descriptive statistics of the full sample of OEFs and ETFs for the years 1992 to 2006. Contrary to Table 1 this table gives 4 snapshots of the sample in 4 different years in order to highlight the time trends in the data. OEFs are Open-Ended Index Mutual Funds. All OEF share classes are collapsed to one entity. ETFs are Exchange Traded Funds tracking indexes and trading on the secondary market. All sizes are in millions of dollars. The percentage growth numbers are winsorized at the 2% level.

Year		1992	1996			2001			2006		
Variable		OEF	OEF	ETF	All	OEF	ETF	All	OEF	ETF	All
Number of Funds		62	81	19	100	182	110	292	200	320	520
Number of Families		30	40	2	42	79	8	86	72	23	92
Number of Indexes		15	21	19	38	32	108	132	55	268	283
Funds per Family	Mean	1.90	2.00	9.50	2.36	2.30	13.75	3.40	2.78	13.87	5.64
	Median	1	1	10	2	1	4	1	1	4	2
	Std	1.81	1.89	10.61	2.96	2.45	24.56	8.17	4.28	25.21	14.50
Number of Indexes per Family	Mean	1.76	1.76	9.50	2.19	2.03	13.75	3.26	2.60	13.70	5.22
	Median	1	1	10	2	1	4	1	1	6	2
	Std	1.55	1.58	10.61	2.96	1.94	24.56	8.45	3.91	23.75	13.03
Fund Size ($M)	Mean	281.84	1,037.22	49.91	849.63	1,655.36	436.00	1,196.02	3,077.35	1,306.63	1,988.99
	Median	88.42	259.64	6.00	189.23	195.59	69.33	121.01	386.59	187.48	268.35
	Std	860.69	3,608.74	159.02	3,267.99	7,315.70	1,718.16	5,894.47	11,536.84	4,756.05	8,109.44
Family Size ($M)	Mean	575.63	2,100.36	474.14	2,022.93	3,813.62	5,995.03	4,060.89	8,548.20	18,122.43	11,220.50
	Median	179.36	317.90	474.14	317.90	431.48	3,086.48	468.67	952.95	1,996.18	1,091.17
	Std	1,780.78	8,731.34	464.77	8,523.24	21,674.23	7,814.93	21,001.50	50,520.57	50,000.55	53,017.65
Percentage Fund Growth	Mean	57.00	64.08	277.31	67.72	7.40	389.11	106.53	27.71	107.78	19.00
	Median	39.17	56.44	277.31	56.44	-3.62	58.08	-0.68	16.51	54.11	19.00
	Std	92.69	56.58	357.10	65.74	43.51	1,026.96	305.29	45.03	150.76	19.00
Percentage Family Growth	Mean	68.06	68.60	39.19	67.71	-3.57	121.63	11.42	16.11	93.30	27.40
	Median	51.00	60.07	39.19	59.30	-7.89	75.31	-4.43	9.81	46.12	15.33
	Std	96.80	71.52	0.00	70.58	23.42	116.30	64.31	31.16	130.15	50.94

Table 3: Summary Statistics of Underlying Indexes Tracked

This table provides descriptive statistics of the indexes tracked by the ETFs and OEFs in the sample. The numbers are averages across all indexes over the entire time of the sample, 1992-2006. The indexes are sampled every quarter and the changes are recorded from one quarter to the other. The liquidity measure is based on Amihud (2002). Amihud's Liquidity Measure (l1) is the absolute value of the return of a stock divided by the absolute value of its price multiplied by its volume. We calculate the liquidity measure of each stock in the index using data downloaded from Joel Hasbrouck's website and described in Hasbrouck (2006) and average it across the index either Equally-Weighted (EW) or Value-Weighted (VW) by the market cap of the stock. We use two industry concentration measures. First, we use the Herfindahl-Hirschman Index using either the Fama-French 10 or 48 industries classifications. We assign to each stock in the index an industry based on the firm's SIC code. Second, we use the industry concentration following Kacperczyk, Sialm, and Zheng (2007)

Variable	OEF	ETF	All
Index Size:			
- Number of Stocks in Index	593.66	325.73	487.41
- Market Cap of Index (in $B)	5656.22	2337.15	4443.42
Index Turnover (over a quarter):			
- Number of Stocks Changing in the Index	13.60	7.47	11.22
- Market Cap Changing in the Index (in $B)	378.40	159.23	299.02
Liquidity of Underlying Stocks:			
- Amihud Liquidity Measure (EW)	0.031	0.074	0.060
- Amihud Liquidity Measure (VW)	0.003	0.022	0.016
Industry Concentration:			
- Herfindahl Index (10 industries)	0.23	0.47	0.32
- Herfindahl Index (48 industries)	0.13	0.32	0.20
- KSZ Industry Concentration	0.09	0.31	0.17