Essays in Asset Management: Mutual Funds and Exchange-traded Funds

Xinrui Zheng

Trinity College

September 29, 2021

University of Cambridge

Judge Business School

This dissertation is submitted for the degree of Doctor of Philosophy

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared below and specified in the text.

Chapter 1 "What Determines an Exchange-traded Fund Launch?" and Chapter 3 "Does the Replication Method Affect ETF Tracking Efficiencies?" are my solo-authored papers. Chapter 2 "The More Things Change, The More They Stay The Same: Why Do Mutual Funds Change Sub-advisors?" is co-authored with my supervisors Prof. Pedro Saffi and Prof. David Chambers, as well as Dr. Julia Arnold.

This thesis is not substantially the same as any work that has already been submitted before for any degree or other qualification at the University of Cambridge or any other institution. It does not exceed the 80,000 word limit for the Cambridge Judge Business School Degree Committee.

Essays in Asset Management:
Mutual Funds and Exchange-traded Funds

Xinrui Zheng

September 29, 2021

Abstract

This dissertation consists of three essays related to fund management, and in particular, mutual funds (MFs) and exchange-traded funds (ETFs).

The first essay studies the decision by an asset manager to launch an exchange-traded fund (ETF). Fund families focus on both revenue generation and cost reduction when making launching decisions, with new ETF launches being driven more by investor demand than past performance. The ETF industry exhibits significant economies of scale and scope, allowing larger families to benefit from specialization while giving smaller families pressure to expand their product line. Competitors tend to follow the asset allocation decisions of the three largest ETF providers, unless when it comes to less liquid or highly concentrated objective markets. Finally, a time-to-event analysis shows that an ETF survives for longer if launched by fund families with larger size and higher fees, and whose initiation is not driven by excessive flows into the family.

The second essay studies the effects of managerial turnover and competition on U.S. sub-advised mutual funds (MFs), using changes of subadvisors by 426 funds from January 1995 to December 2016. Sub-advised MFs make turnover decisions based on return-chasing behavior, but these changes neither improve subsequent fund returns and risk measures, nor increase future flows into the fund. Using sub-advisor turnover to change the degree of competition among sub-advisors does not affect the performance of incumbent sub-advisors. Overall, there is no evidence that sub-advisor selection decisions by fund families benefit sub-advised MF's performance. Outperforming sub-advisors with larger style drift are less likely to be hired, and the more a sub-advisor deviates from its investment mandate, the more likely it is to be fired.

The third essay uses 2,290 European equity and fixed income ETFs and studies how the replication method affects the tracking efficiencies of ETFs, especially during market crises. Throughout the 20-year sample period 2001 to 2020, there is no persistent evidence suggesting superior tracking performance of synthetic ETFs. I identify 119 benchmarks followed by both physical and synthetic ETFs simultaneously, and conduct a difference-in-difference analysis around Lehman Brothers bankruptcy, sovereign debt crisis and COVID-19 outbreak. Synthetic ETFs face steeper declines in tracking efficiencies following a sudden increase in counterparty risk, while they are shielded from liquidity shocks. There is a remarkable drop in tracking performance sensitivity to market distress post the global financial crisis.

To my beloved family and friends

Acknowledgements

I am grateful to my supervisors, Prof. Pedro Saffi and Prof. David Chambers, for the guidance, encouragement and opportunities they provided throughout my PhD journey, as well as the support both academically and mentally over the pandemic period.

I would also like to thank Prof. Andrei Kirilenko, Dr. Oğuzhan Karakaş, Prof. Elroy Dimson, Prof. Bart Lambrecht, Prof. Raghavendra Rau and all the Finance Faculty at the Judge Business School for their insights and comments on my work as well as guidance on an academic career. I thank Prof. Marc L. Lipson at Darden School of Business, University of Virginia for providing valuable advice and reference. I thank Dr. Julia Arnold for collecting the data for Chapter 2 "The More Things Change, The More They Stay The Same: Why Do Mutual Funds Change Sub-advisors?". I also thank all the participants at the SFA Annual Meeting (November 2021), European Investment Forum (September 2021), World Finance Conference (December 2020), CJBS Winter Doctoral Conference (December 2020), CERF seminar and CERF cavalcade (May 2020) for insights and comments. I gratefully acknowledge full financial support from Cambridge Endowment for Research in Finance (CERF).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis consists of three essays in asset management related to mutual funds (MFs) and exchange-traded funds (ETFs). The first essay sheds light on the industrial organization of ETFs by examining family decisions to make new ETF offerings. The second essay reflects on the effects of managerial turnover and competition by studying subadvisor turnover decisions of sub-advised MFs. The third essay investigates the effect of the replication method on ETF tracking efficiencies.

The ETF industry has grown enormously since the launch of the SPDR fund by State Street in 1993. The assets under management (AUM) of US-based ETFs alone surpassed 5 trillion USD by November 2020 , and ETFs now comprise nearly a third of the trading activities on the US stock market. Competition and innovation in the ETF marketplace are encouraged by the regulators. However, the ETF market is highly concentrated, with over 80% of the total net assets (TNAs) controlled by the "Big Three" asset managers: BlackRock, Vanguard, and State Street. This raises regulatory concerns over the healthiness of competition in the ETF marketplace.

The first essay addresses four main questions. First, what are the reasons behind ETF launches? Second, how do the determinants of ETF launches compare with those for open-ended MFs? Third, how does the presence of the Big Three shape the ETF industry? Fourth, can we predict the lifespan of an ETF based on characteristics of the family and the ETF's investment objective upon inception? Findings from this essay suggest that the decisions to launch ETFs and open-ended MFs differ in important ways. While the active fund families rely heavily on past performance in making new launches, ETF families pay more attention to matching investor demand by chasing flows and trading volumes in the market. ETF families tend to attract market makers for their newly-launched product by leaving greater arbitrage profits through the unique in-kind creation and redemption mechanism. Families making new ETF launches are cost-conscious in both limiting the trading cost and reducing the per unit cost by exploiting the economies of scale. The Big Three tend to lead the asset allocation decisions of smaller families. However, competitors are less likely to follow the leader into less liquid or highly concentrated objective markets. Finally, the lifespan of a newly-launched ETF can be predicted by the family and objective-level characteristics upon inception. A time-to-event analysis shows that an ETF survives for longer if launched by fund families with larger size and higher fees, and whose initiation is not driven by excessive flows into the family.

Mutual fund families often delegate or outsource the asset management of a mutual fund to sub-advisors rather than managing the assets in-house. Given that mutual funds that outsource management are typically more sophisticated, knowledgeable, and better resourced than the average mutual fund investor, one might expect fund families to do a better job in picking outperforming managers. The large number of sub-advisor hiring and firing decisions made by fund families allows us to test this hypothesis. Between January 1995 and December 2016, we count 1,239 hiring and 809 firing decisions made by fund families, across 426 sub-advised mutual funds. Fund families may choose to replace a sub-advisor for "good" reasons, such as to improve fund returns and to lower risk, or for "bad" ones, such as to increase fund flows and asset management revenues despite any improvement in performance.

Results from the second essay show that advisors exhibit similar behaviour to retail investors when selecting a mutual fund (Sirri and Tufano (1998)). Fund families tend to fire sub-advisors following underperformance and hire them after outperformance. When the fund itself is outperforming, there are fewer hirings and firings. Furthermore, such changes do not lead to higher returns, lower volatility, better Sharpe ratios, and higher asset flows. We also do not find that competition among sub-advisors improve performance. Turnover does not affect the risk-adjusted returns of incumbent sub-advisors, and their subsequent performance is not different from those of recently hired/fired sub-advisors. Finally, the likelihood that a sub-advisor is fired increases with its degree of style drift (measured by differences of factor loadings relative to its Morningstar style's average) and that the likelihood of being hired following outperformance is reduced by the style drift.

There are two fundamentally different ways that exchange-traded funds (ETFs) can replicate their underlying benchmark indices, namely physical replication and synthetic replication.[1] Physical replication involves holding all constituent securities or a representative sample of the benchmark index. Synthetic replication achieves the benchmark return by entering into a total return swap or other derivative contract with a counterparty, typically a large investment bank. The first synthetic ETF was introduced on the French market in 2001. Since then, synthetic structures have become more popular in Europe than in the US due to different regulations on fund's legal structures.[2] In this paper, I examine all equity and fixed income ETFs in European from 01 January 2001 to 31 December 2020 to see if the replication method affects the tracking error. After the global financial crisis, synthetic ETFs has been widely criticized by regulators and financial advisors for their complexity, lack of transparency and counterparty risk. However, there is a resurgence in interest for synthetic ETFs, particularly those providing exposure to the US equity market due to their tax advantage over their physical peers (Zarate et al., 2021).[3] It is therefore important to known whether fund families offering synthetic ETFs learn from market failures and improve the risk management of synthetic structures since then.

The third essay addresses the concern by answering the following three questions. First, do synthetic ETFs posses superior tracking ability compared to physically-replicated ones? Second, which replication method can better withstand market distress? Third, is there any

[1]Physical replication can be divided into full replication and sampling, synthetic replication can be based on either total return swaps or other derivatives, such as futures contracts. A more detailed classification can be found at: http://www.argos-tsp.com/en/research/argos-finneo/terminological-research/summary-of-index-replication-methods-used-by-etf-providers.html.

[2]Most US-registered ETFs are governed by the Investment Company Act 1940 (ICA), which prohibits transactions between fund and its affiliate as well as other forms of self-dealing. As a result, organizing an ETF using synthetic structure becomes complicated. On the other hand, the majority of European-listed ETFs are regulated by UCITS, which allows the use of exchange-traded as well as OTC derivatives to achieve investment objectives, and therefore synthetic replication becomes more popular in Europe.

[3]BlackRock, once being a major critic of the synthetic structure, has launched a swap-based S&P 500 UCITS ETF in September 2020. Here is the Financial Times article: https://www.ft.com/content/6600bd7f-5433-47d3-a2df-04411e6de75b.

improvement on risk management after the global financial crisis, especially in terms of the swap counterparty risk of synthetic ETFs? I find no evidence of persistent superior tracking ability of synthetic ETFs across the sample period, especially after controlling for heterogeneity across the investment objectives. There are significant cross-sectional variations in tracking errors. Furthermore, after the global financial crisis, I observe a large reduction in tracking errors. Synthetic ETFs face steeper declines in tracking efficiencies after a sudden increase in counterparty risk. But during liquidity shocks, their tracking ability is less affected relative to the physical ones. I explain the relative tracking performance between physical and synthetic replication around market crisis by describing the trade-off between counterparty and liquidity risk that dominates the market. Finally, I find that the tracking performance of both physical and synthetic ETFs becomes significantly less sensitive to market distress after the global financial crisis. Specifically, synthetic equity ETFs demonstrate superior tracking ability in terms of both lower tracking errors and lower sensitivity to market turbulence in the post-crisis sample period.

The rest of this thesis is organized as follows. The next three chapters present my three essays, followed by a chapter concluding with my main findings and implications. Several suggestions on future research are also included at the end of the last chapter.

Chapter 2

What Determines an Exchange-traded Fund Launch?

Xinrui Zheng*

September 01, 2020

Abstract

This paper studies the decision by an asset manager to launch an exchange-traded fund (ETF). Fund families focus on both revenue generation and cost reduction when making launching decisions, with new ETF launches being driven more by investor demand than past performance. The ETF industry exhibits significant economies of scale and scope, allowing larger families to benefit from specialization while giving smaller families pressure to expand their product line. Competitors tend to follow the asset allocation decisions of the three largest ETF providers, unless when it comes to less liquid or highly concentrated objective markets. Finally, a time-to-event analysis shows that an ETF survives for longer if launched by fund families with larger size and higher fees, and whose initiation is not driven by excessive flows into the family.

Keywords: Exchange-traded fund (ETF), launching decision, fund family, investor demand, flow, volume, market concentration
JEL classification: G10, G11, G12

*Judge Business School, University of Cambridge, Address: Trinity College, Cambridge, UK CB2 1TQ, email: xz359@cam.ac.uk. I thank Pedro Saffi, Marc Lipson, Andrei Kirilenko, David Chambers and all the participants at the SFA Annual Meeting (November 2021), World Finance Conference (December 2020), CJBS Winter Doctoral Conference (December 2020), CERF seminar and CERF cavalcade (May 2020) for insights and comments. I gratefully acknowledge full financial support from Cambridge Endowment for Research in Finance (CERF). All remaining errors are my own.

2.1 Introduction

The exchange-traded fund (ETF) industry has grown enormously since the launch of the SPDR fund by State Street in 1993. The assets under management (AUM) of US-based ETFs alone surpassed 5 trillion USD by November 2020 , and ETFs now comprise nearly a third of the trading activities on the US stock market.[1] Figure 2.1 shows that new ETF launches number from several dozen to hundreds in each year over the past decade, and the aggregate AUM of the ETF industry has grown exponentially. Competition and innovation in the ETF marketplace are encouraged by the regulators. For instance, in September 2019 the Securities and Exchange Commission (SEC) adopted a new rule to facilitate ETF market entrants by effectively watering down the 'exemptive relief' requirements.[2]

One distinguishing feature of ETFs relative to open-ended mutual funds (MFs) is the extreme market concentration, which raises regulatory concerns that it may stifle competition.[3] Figure 2.2 shows that over 80% of the total net assets (TNAs) in ETFs are managed by one of the "Big Three" asset managers: BlackRock, Vanguard, and State Street. With its exponential growth and popularity among financial advisors, the ETF industry has undoubtedly been placed under the spotlight by investors and regulators alike. However, academic research on the industrial organization of ETFs is limited. The decision of fund families to launch new ETFs has not been academically studied before.[4]

This paper therefore fills in this gap by addressing four main questions. First, what are the reasons behind ETF launches? Second, how do the determinants of ETF launches compare

[1]According to a survey by the Financial Planning Association and the Journal of Financial Planning in April 2019 with 392 respondents, ETFs are the most popular investment vehicle and the share of financial advisors recommending them to their clients increased from 44 percent in 2008 to 88 percent in 2019. Source link: https://www.financialplanningassociation.org/business-success/ResearchandPracticeInstitute/Documents/2019-Trends_%20in_Investing_Report.pdf

[2]The rule aims to streamline the conditions around exemptive relief, which has been costly and time-consuming to new ETF providers. The press release and fact sheet are available on the SEC website: https://www.sec.gov/news/press-release/2019-190

[3]Barron's reported in April, 2019 the concern from the SEC in an online article: https://www.barrons.com/articles/etfs-are-dominated-by-blackrock-vanguard-and-state-street-the-sec-is-concerned-51554512133?mod=hp_DAY_2.

[4]Sherrill and Stark (2018) studied the determinants on ETF liquidation. Ben-David et al. (2021) focused on the product differentiation perspective in the innovation of specialized ETFs. To my best knowledge, there is no academic paper which systematically studies the launching decisions of ETFs.

with those for open-ended MFs? Third, how does the presence of the Big Three shape the ETF industry? Fourth, can we predict the lifespan of an ETF based on characteristics of the family and the ETF's investment objective upon inception? The main findings of this paper suggest that the decisions to launch ETFs and open-ended MFs differ in important ways. While the active fund families rely heavily on past performance in making new launches, ETF families pay more attention to matching investor demand by chasing flows and trading volumes in the market. ETF families tend to attract market makers for their newly-launched product by leaving greater arbitrage profits through the unique in-kind creation and redemption mechanism. Families making new ETF launches are cost-conscious in both limiting the trading cost and reducing the per unit cost by exploiting the economies of scale. The Big Three tend to lead the asset allocation decisions of smaller families. However, competitors are less likely to follow the leader into less liquid or highly concentrated objective markets. Finally, the lifespan of a newly-launched ETF can be predicted by the family and objective-level characteristics upon inception. A time-to-event analysis shows that an ETF survives for longer if launched by fund families with larger size and higher fees, and whose initiation is not driven by excessive flows into the family.

It is not surprising to find that fund families make launching decisions according to investor demand, however the difference in the effect of investor demand on fund offerings between ETFs and open-ended MFs is more subtle. Actively managed MF managers may limit the size of funds and thus the value to the fund family, as the ability to implement market-beating investment strategies decreases with fund size (Berk and Green, 2004; Pástor et al., 2015). Conversely, ETFs are mostly passive vehicles, with their value to the fund family being driven more directly by the volume of investors. Khorana and Servaes (1999) study the determinants of MF starts. They find that launch decisions of the opend-ended MFs are related to three broad sets of factors: the ability to generate additional fee income, economies of scale, and follow-the-leader strategy. In this paper, I examine the intuitive question of whether common characteristics shared between ETFs and open-ended MFs affect their launches in the same way. Furthermore, structural differences between ETFs and open-ended MFs indicate the potential existence of additional factors in determining ETF inceptions that are not present in the MF literature. Unlike the open-ended funds, ETFs are traded on stock exchanges and enjoy intra-day liquidity. I examine stock exchange characteristics that are unique to ETFs to study the effect of market liquidity on ETF launching decisions. Given the extensive debate on active versus passive management (Levy and Lieberman, 2016; Garleanu and Pedersen, 2019), this paper explores whether family decision to launch new ETFs in a particular investment objective is correlated with characteristics of the MF equivalent.

I first show that ETF launching decisions are affected by families' desire to generate incremental profits, but through different channels relative to open-ended MFs. Prior literature documents the predominant role played by past performance at both the family-level and the investment objective-level in determining the initiation and termination of open-ended MFs (Khorana and Servaes, 1999; Brown and Goetzmann, 1995; Lunde et al., 1999). Superior abnormal returns attract investor flows (Ippolito, 1992; Sirri and Tufano, 1998) and hence facilitate growth of the family, justifying the emphasis on past performance by MF families.[5] However, the majority of ETFs are still passive index trackers with no explicit goal of generating alpha.[6] This rationalizes ETF families' focus on flows, total AUM, and aggregate fees to gain a competitive edge. I find empirical evidence suggesting that ETF families are more inclined to profit from flow and volume. There is no significant correlation though between ETF launches and past performance, either at the family-level or the investment objective-level. Instead, I find a positive and significant relationship between the likelihood of ETF launches and the prior-12-month dollar volume in the fund family and in the ETF's investment objective. Similar pattern has been found on the prior-12-month net flow into the investment objective. This contrasts with the findings for open-ended MFs, where the effect of flows becomes trivial after controlling for past performance (Khorana and Servaes, 1999).

New ETF launches can be compared to the initial public offerings (IPOs). Loughran and Ritter (2002) show that issuers do not get upset about leaving significant amount of money on the table during IPOs and argue that it may serve as a form of indirect compensation to the

[5]Furthermore, an increasing proportion of mutual funds are adopting an incentive fee structure that rewards the abnormal returns directly (Elton et al., 2003). It may also help explain the heavy reliance of MF initiations on past performance, which are absent on the ETF sample.

[6]Despite the recent emergence of actively managed ETFs into the marketplace, they still represent less than 10% of the overall AUM.

underwriters. Inspired by the IPO literature, I argue that fund families are willing to leave arbitrage opportunities through the in-kind creation and redemption process when launching new ETFs as they expect to attract more participation from the market makers, i.e. the authorized participants (APs). The in-kind creation and redemption mechanism is an important structural difference that distinguishes ETFs from open-ended MFs (Antoniewicz and Heinrichs, 2014). It takes place on a daily basis and ensures the value of the ETF shares are kept close to the value of the underlying basket of securities (Ben-David et al., 2018). In this paper, I find empirical evidence suggesting that families are more likely to launch in objective markets with more arbitrage opportunities, measured by higher tracking errors, to the market makers.

Like open-ended MFs, fees charged by ETFs consist of a fixed percentage of its total net asset (TNA). Therefore, to maximize their fee income, families desire both a larger asset base and a higher percentage fee. To ensure a substantial asset base for the ETF offering, the potential size of the market is an important consideration for family launching decisions. Moreover, since investors may switch between the active and passive alternatives with the same investment objective (Garleanu and Pedersen, 2019), the aggregate TNA of the actively managed open-ended MFs in the given objective may serve as a proxy for the customer base and hence the growth potential of the objective market. I find that new ETFs are more likely to be launched in objectives with larger TNAs and in those experiencing larger inflows, supporting the argument that families care about the growth potential of the objective market. Additionally, ETFs are more likely to be launched in objectives where the aggregate TNA of the MF equivalent is larger. Families with higher expense ratios are also more likely to launch new ETFs. However, at the objective level, the impact of fees is less clear. On the one hand, fee competition among ETFs is widely reported.[7] Issuers may compete for cost-conscious investors by launching ETFs in objectives with lower expense ratios. On the other hand, as competition increases, fund families start to compete through an alternative channel, using specialized ETFs that track niche markets and can charge higher fees (Ben-David et al., 2021).

In addition to revenue growth, profit maximization can also be achieved via cost reduc-

[7]The following Bloomberg article in March 2019 is an example. `https://www.bloomberg.com/opinion/articles/2019-03-22/etf-fee-wars-are-no-laughing-matter.`

tion. Fund families face both fixed costs and variable costs in running an ETF. While fixed costs include expenditure on R&D, marketing expenses and regulatory charges, variable costs mainly stem from transaction costs triggered by portfolio turnover. In the realm of passive ETFs, reconstitution of the benchmark index is the main driver of portfolio turnover. From the variable costs' perspective, we expect that fund families are more likely to launch in objectives that are cheaper to operate. Consistently, I find that new ETFs are more likely to be launched in objectives with lower portfolio turnover.

Another way to achieve cost reduction is by exploiting economies of scale and scope to effectively share the fixed costs of running an ETF. Empirical evidence suggests that ETFs are more likely to be launched by larger families and families who have launched in the prior year. Notably, I find supporting evidence to the argument that equity ETFs benefit from specialization that outweighs the potential cost of cannibalization. Under a panel logistic regression framework, I find that families are more likely to launch in an objective if it represents a greater proportion of the family's TNA. This contradicts the results for equity MFs by Khorana and Servaes (1999) but is consistent with Evans (2010).[8] Meanwhile, in the ETF industry, the benefits of scope are valued as much as scale. Many retail brokerages, pension plans, and online platforms are reducing the number of fund providers they offer to their clients. Therefore, asset managers offering the full spectrum of products will gain a competitive edge. To understand the choice by families to launch an ETF in a new objective versus one in which the family already offers a product, I estimate a multinomial logit regression to examine each decision relative to no launching. I find that smaller families dominate the decision to launch in objectives where the family has no prior investment. These families face more pressure to grow and expand to achieve a larger scope needed to compete in the ETF industry.

To address my third research question, I investigate how the presence of the Big Three affects competition among new market entrants. Figure 2 shows the market share of the ETF industry controlled by the Big Three from 1996 to 2018, along with the number of new ETF

[8] Khorana and Servaes (1999) find empirical evidence on specialization in bond MFs but not in equity MFs. In a later study on equity MFs, Evans (2010) find that families with a large percentage of assets invested in a given investment objective are more likely to launch an additional fund in that investment objective, consistent with a desire to specialize.

launches by families within and outside the Big Three. Though the graph indicates that the number of new entrants from non-Big Three families has increased dramatically through years, regulators are still concerned about whether the extreme market concentration in the ETF industry reduces competition. The Big Three enjoying better brand recognition and a wider customer base, attract more flows and provide higher liquidity. These competitive advantages allow them to develop more effectively and become the first-movers in most investment objectives. By following asset allocation decisions of the Big Three, other families can save substantially on research costs and customer development. On the other hand, with the scale and scope economies favoring the Big Three and switching costs being substantial within less liquid investment objectives, the first mover advantage is more likely to persist. Therefore, additional barriers are imposed on the new market entrants into such objectives. I find that non-Big Three families tend to make new ETF launches in objectives where the Big Three has launched in the prior year, which is consistent with the findings for open-ended MFs (Khorana and Servaes, 1999). However, to the extent that the market for an investment objective is less liquid or dominated by the Big Three, the follow-the-leader behaviour among non-Big Three families is diminished. As a robustness check, I test if the family and objective-level characteristics affect launching decisions of the Big Three differently relative to the non-Big Three. Empirical pattern suggests that the impact of most characteristics is in the same direction for both the Big Three and non-Big Three, though the economic significance may vary.

Finally, with respect to the fourth research question, I conduct a time-to-event analysis to study the survival time of ETF offerings. There are significant costs in launching an ETF. It typically takes between \$750,000 and \$1.25 million to get through the exemptive relief, the prospectus and all the contracts, besides costs associated with research, marketing and maintenance. Therefore, a fund needs to grow rapidly and attract sufficient inflows to remain profitable in the long run. I find that the lifespan of an ETF offering can be predicted from family and objective-level characteristics upon inception. ETFs launched by larger families and families charging higher fees are more likely to survive for longer. In the presence of substantial scale economies, it is easier for larger families to attract flows and save on the per unit cost. Besides, families charging higher fees can maintain a higher profit margin. However, I find that ETFs launched following large inflows into the fund family are more likely to fail at an earlier age. One possible explanation is that excessive inflows into the fund family may evoke managerial hubris as proposed in the corporate takeover literature (Roll, 1986). At the objective-level, ETFs launched in objectives with higher expense ratios are more likely to fail earlier. This is consistent with the notion that most ETFs are passive vehicles tracking broad-market indices and the ETF industry is extremely competitive on fees. Moreover, it is also related to findings by Ben-David et al. (2021) that expensive specialized ETFs perform poorly post-launching. Finally, previous literature on MFs documents that launching a new fund into "hot" or "trendy" objectives may lead to future underperformance (Greene and Stark, 2016). Similarly, I find that ETFs who are launched due to large trading activities in the objective market are more likely to fail at a younger age.

This paper carries out a systematic study over the determinants on ETF initiations, contributing to the literature from several perspectives. First, I build upon the literature concerning the industrial organization of the open-ended funds to show how the decision to launch an ETF is affected by fund characteristics in distinct ways relative to open-ended MFs. Second, I provide evidence on how market conditions, including liquidity and market concentration, affect the competition and growth of the ETF industry. Prior literature has documented the effect of market quality on ETF flows (Clifford et al., 2014). In this paper, I discuss a more direct channel of the industry expansion, namely the emergence of new market entrants.

The rest of this paper is organized as follows. Section 2 lists the major hypotheses. Section 3 describes the data and shows some summary statistics. Section 4 explains the research design and elaborates the empirical results from the analysis. Section 5 contains the robustness check. And Section 6 concludes.

2.2 Hypotheses

Our purpose is to examine the determinants of ETF launches by fund families and how they compare to open-ended MFs. For both ETFs and open-ended MFs, the launching decision is rational only if the expected benefits outweigh the associated costs. Families benefit from additional fee income, which consists of a percentage of the total AUM. Therefore, expanding the asset base and charging a higher percentage fee are of the best interest to both types of funds. However, the channels through which families generate profits can be different. While active funds rely on superior past performance to advertise managerial skills and attract investments, most ETFs are passively managed tracking their benchmark index. I hence expect no significant correlation between ETF launching decisions and the abnormal returns on either the family level or the investment objective level. Instead, I posit that family decisions to launch new ETFs are more likely to be driven by investor demand. The growth potential of an objective market may be signalled by prior flows and trading volume. Moreover, families who have attracted large inflows may enjoy the "halo effect" and flows are more likely to spill over to the newly launched product. This leads to the first major hypothesis:

Hypothesis 1 *The likelihood of an ETF launch is positively related to prior net flows and trading volumes in the family and the investment objective.*

While the above arguments speak from the demand side of an ETF, the launch of a fund also depends on supply-side characteristics. ETF shares are created and redeemed through in-kind transactions between the ETF providers and the authorized participants (APs) on the primary market. As market makers, the APs are essential for the distribution and marketing of the ETF products. They make profits through the bid-ask spread, smart management of their securities inventory, and through arbitrage activities. In an efficient market, when price discrepancies between the ETF shares and the underlying basket emerge, the APs may buy whichever is cheaper on the secondary market in exchange for the other with the ETF provider on the primary market. In this way, the APs are able to earn arbitrage profits and effectively drive the two prices to convergence. Higher tracking error in a given objective market leaves more room for arbitrage activities by the APs. Therefore, in the same spirit as the IPO firms leave money on the table as indirect compensation to the underwriters (Loughran and Ritter, 2002), I posit that fund families deliberately leave arbitrage opportunities to attract participation of the APs. This leads to the following hypothesis:

Hypothesis 2 *The likelihood of new ETF launches in a given investment objective is positively* *related to the average tracking error of all funds in the objective.*

Similar to the market for open-ended MFs, there exist substantial scale and scope economies in the ETF market. I hence posit that larger families, families offering a wider range of products, and families who have launched in the prior year are more likely to launch a new ETF. Smaller families facing the curse to "grow-or-die" are more likely to expand their product line to withstand competitions in the earlier stage of their lives, while more established families may benefit from specialization and the associated economies of scale. This leads to the next major hypothesis:

Hypothesis 3 *The decision to launch a new ETF in an objective with no prior investment within the family is dominated by smaller-sized families or families with smaller scope. The likelihood of launching a new ETF in an existing objective is positively related to the percentage of family asset invested in the corresponding investment objective.*

The ETF industry is highly concentrated, with over 80% of the aggregate AUM controlled by the Big Three. The presence of significant scale and scope economies favours the Big Three and allows them to become the first innovators. However, to the extent that non-Big Three families could replicate the innovation without incurring significant costs and benefit from the overall growth of the market, it is rational for them to follow the leader and take the second mover advantage (Reinganum, 1985). Hence, similar to the behaviour of MF families, I posit that an ETF launched in a given objective by the Big Three during the previous year would increase the likelihood of non-Big Three families to launch in the same objective. However, the profitability of the second mover is related to factors such as brand loyalty and switching cost in the market. For individual investors, moving from one fund to another triggers capital gain tax. Institutional investors often suffer from inertia. For illiquid asset, the spread could also bring in substantial trading cost. Therefore, I take a closer look at the impact of the Big Three on the competition of other families. Though the follow-the-leader strategy found in the MF literature still exists in the ETF sample, I posit that non-Big Three families would hesitate in entering investment objectives dominated by the Big Three. Also, in objectives with less liquidity and thus higher switching cost, the first-mover advantage of the Big Three are more likely to sustain, which poses additional barrier on new market entrants. Here follows the next major hypothesis:

Hypothesis 4 *The likelihood of non-Big Three families to launch new ETFs in an objective where the Big Three has launched in the previous year is negatively related to the market share of the objective controlled by the Big Three and the average spread of the objective.*

Given the inevitable costs to launch an ETF, the fund needs to attract sufficient inflows and grow rapidly to remain profitable. I argue that family and objective characteristics upon inception have significant predicting power on the lifespan of ETF offerings. At the family level, size can be an important determinant. It is generally harder for smaller families to attract flows due to the disadvantage on brand recognition and customer base, etc. Also, families charging higher fees are able to maintain a higher profit margin. However, families that experienced excessive inflows in the prior year may be affected by hubris beliefs as proposed in the corporate takeover literature (Roll, 1986). At the objective level, the ETF industry is extremely competitive on expense ratios due to the passive nature. Hence, ETF products launched in lower-fee objectives possess a competitive advantage on this end. Moreover, previous literature on MFs documented that launching a new fund into hot or trendy objectives may lead to future underperformance (Greene and Stark, 2016). In a similar vein, I argue that ETFs who are launched due to large trading activities in the objective market are more likely to fail in a younger age. Here comes the last major hypothesis:

Hypothesis 5 *ETFs launched by larger and higher-fee families, and whose initiations are not driven by capital flows into the families are more likely to survive for longer. ETFs launched in lower-fee objectives and whose launching decision is not affected by large trading activities in the objective market are more likely to live a longer life.*

In Table A2.1, I summarize all the ex-ante predictions on how various family and objective-level characteristics may affect the likelihood of new ETF launches. The hypotheses are divided into three broad categories, namely profit maximization, scale and scope economies and the impact of the Big Three. Predictions are posed in either the whole sample or two different subsamples of the existing families and the non-Big Three families where appropriate.

2.3 Data and Sample Descriptions

In this study, I mainly focus on 1,859 US equity ETFs from January 1996 to December 2018, taken from Morningstar Direct Database. Both surviving and delisted funds are included in the sample to avoid survivorship bias.[9] During the sample period, I observe 1,756 ETF launches across 155 fund families and 74 Morningstar Institutional Categories. I obtain monthly observations on the gross and net returns, as well as the total net assets (TNAs). Fund flows are then calculated as the change in TNAs net of any return effect, i.e.

$$Net\ Flow_{i,t} = TNA_{i,t} - TNA_{i,t-1} * (1 + R_{i,t}), \tag{2.1}$$

where $TNA_{i,t}$ denotes the total net assets in fund i at the end of month t, and $R_{i,t}$ is the gross return of fund i during month t. I also obtain daily observations on the prices, trading volumes and bid-ask spreads. For the daily variables, monthly averages are computed to fit into the fund-month panel. Taking the dollar volume for instance:

$$Dollar\ Volume_{i,t} = \frac{1}{T}\sum_{t=1}^{T} Daily\ Trading\ Price_{i,t} * Daily\ Volume_{i,t}, \tag{2.2}$$

where T denotes the number of trading days within month t. Finally, expense ratios and portfolio turnovers are reported annually on Morningstar. The annual report net expense ratio represents the percentage of fund TNA used to pay for operating expenses, management fees, administrative fees, and all other asset-based costs incurred. The portfolio turnover is

[9] I begin with 3,305 exchange-traded products (ETPs) whose primary share class is listed in U.S. exchanges, then I exclude 359 exchange-traded notes (ETNs) and exchange-traded commodities (ETCs) from the original sample. Morningstar uses "ETF" as an umbrella term to refer to a range of different ETPs, including ETFs, ETNs and ETCs. Unlike ETFs, ETNs and ETCs are more accurately described as debt securities that are less relevant to our study.

defined as

$$Portfolio\ Turnover_{i,t} = \frac{\min(Purchase_{i,t}, Sale_{i,t})}{TNA_{i,t}}, \tag{2.3}$$

the numerator denotes the total amount of either purchase or sale by fund i during year t whichever is less. Family and objective-level measures are obtained by aggregating all ETFs within the family or objective into one portfolio and taking the equally-weighted average. The value-weighted averages are also computed and included in the robustness checks.

To examine characteristics of the MF equivalent, I further obtain monthly observations on returns, TNAs and expense ratios for 9,386 US equity MFs during the same period. Finally, monthly returns on the benchmark indices are collected from Morningstar. Benchmark-adjusted abnormal returns of the funds are then calculated as the difference between the realized returns and the predicted returns with beta estimates derived from the following regression using a prior 36-month rolling window:

$$R_{i,t} - r_f = \alpha_i + \beta_{i,t}(R_{BMK_i,t} - r_f) + \varepsilon_{i,t}, \tag{2.4}$$

where $(R_{BMK_i,t} - r_f)$ denotes the excess return of the primary benchmark over the risk-free rate for fund i during month t. For fund observations whose primary benchmark is missing, the default benchmark is obtained by taking the most commonly used (sorted by the aggregate TNA of funds following the index) primary benchmark for ETFs in each Morningstar Institutional Category, as shown in Table A2.2.

Table 2.1 reports the descriptive statistics of the fund-month observations on major fund characteristics, including net flow, dollar volume, tracking error, expense ratio, portfolio turnover, TNA, family TNA, bid-ask spread and benchmark-adjusted return. In panel A, summary statistics are reported for all funds in the sample.[10] Panel B compares between ETFs from the Big Three versus non-Big Three families. Panel C compares between active versus passive ETFs. The "MeanDiff" column in Panel B and C reports the difference in the sample means, together with the statistical significance from a two-sided t-test. Out of the 1,859 US equity

[10]Notice that the minimum expense ratio is negative, this is due to security lending activities by the ETFs.

ETFs, 505 ETFs are launched by the three biggest families, while 1,354 ETFs are launched by 152 non-Big Three families. The line plot in Figure 2.2 shows that the Big Three controlled over 80% of the total AUM in each year over the past decade. Panel B indicates that funds launched by the Big Three are generally larger in size and attract more flows and trading volumes, hence they enjoy better liquidity with lower bid-ask spread. They also charge lower fees and track their benchmarks more closely. Out of the whole sample, 151 US equity ETFs are actively managed, representing around 8% in numbers and less than 0.5% in terms of the AUM by the end of the sample period.[11] Panel C shows that the size of passive ETFs are on average larger than the active ETFs, and they are more likely to be launched by larger families, charging lower fees. Passive ETFs tend to attract more flows and trading volumes and hence provide better liquidity with lower bid-ask spread. There are generally less portfolio re-balancing in passive ETFs and hence lead to lower portfolio turnovers and higher tracking errors.

2.4 Results

2.4.1 General Factors Affecting ETF Launching Decisions

In this section, I investigate the overall determinants of an ETF launch and test Hypothesis 1 and 2 from both the demand side and the supply side of a fund. On the demand side, I explore the differential effect of investor demand and past performance on the launching decisions of ETFs versus opend-ended MFs. On the supply side, I examine whether the launching families care about participation from the market makers, i.e. the APs. In order to do so, I use the

[11] By December 2018, the total AUM of all actively-managed US equity ETFs is 9.74 billion USD, as apposed to 2.61 trillion USD managed by passive ETFs.

following panel logistic framework:

$$\begin{aligned}\log\left(\frac{p_{ij,t}}{1-p_{ij,t}}\right) = \quad & \beta_0 + \beta_1 Objective\ Flow_{j,t-1} + \beta_2 Family\ Flow_{i,t-1} \\ & + \beta_3 Objective\ Dollar\ Volume_{j,t-1} + \beta_4 Family\ Dollar\ Volume_{i,t-1} \\ & + \beta_5 Objective\ Tracking\ Error_{j,t-1} + \textstyle\sum_{k=1}^{n} \beta_k X_{ij,t} + \gamma_t + \varepsilon_{ij,t},\end{aligned} \tag{2.5}$$

where $p_{ij,t}$ denotes the probability of family i launching a new ETF in objective j during month t. The matrix $X_{ij,t}$ includes control variables on both the family level and the investment objective level. In order to make clear contrast between active and passive investments, all the actively managed ETFs and indexed MFs are excluded from this regression.

Table 2.2 presents the coefficient estimates together with the marginal effects of the panel logistic regression using all family observations. In Model (1), I include in explanatory variables only the measures on flows, volumes, returns and tracking errors as discussed in Hypothesis 1 and 2. As a flow measure, I include the prior 12-month average flow rank as in (Evans, 2010). A fractional rank between zero and one is assigned to each family or investment objective in each month according to the net flow. By using the flow rank instead of the percentage flow, the regression is immune from the distortion by the outliers and the undesired market turbulence through time. As a measure of liquidity, I include the prior-12-month average rank of the dollar trading volumes, which is constructed in the same manner as the flow ranks. To measure the arbitrage potential in an objective market, I include the mean tracking error of all ETFs within the same investment objective, calculated as the standard deviation of the prior-12-month excess returns over their primary benchmarks. The default objective benchmark (as listed in Table A2.2) is assigned when the primary benchmark is missing. As for the performance measure, I use the prior-12-month average benchmark-adjusted returns of the family and investment objective. Unlike the actively managed funds, for whom the Carhart four-factor alpha (Carhart, 1997) may be a better indicator of the absolute performance. In the realm of passive ETFs, it makes better sense to compare the performance of a fund with the stated benchmark index.

In Model (2), I also include the expense ratio of the family/objective in the prior year as a proxy for the percentage fee charged. Portfolio turnover of the family/objective in the prior year is included to account for cost reduction considerations. Finally, to explore the scale economies as well as to control for the well-documented relationship between flow and fund size, I include the logarithm of lagged family and objective size. In Model (3), I explore the effect of economies of scope on ETF launches. Family size is replaced by the number of objectives covered in the family.[12] In Model (4), I consider several objective level characteristics of open-ended MFs, given the documented competition on investor flows between active and passive funds (Levy and Lieberman, 2016; Garleanu and Pedersen, 2019). I substitute the objective flow, size, expense ratio and performance measures with the corresponding measures for the open-ended MF equivalent. More detailed definitions on the variables can be found in the appendix Table A2.3.

I make the assumption that the ETF launching decisions are independent across different families, though decisions made by the same family are more likely to be correlated. Therefore, the standard errors are clustered by fund family. Calendar year fixed effects are included. In the columns "Marginal Effect", I report the annualized percentage change in the probability of an ETF launch when each explanatory variable is increased by one standard deviation and all other variables are set equal to the mean of zero. For indicator variables, this represents the percentage change in probability when the indicator variable increases from zero to one.

Hypotheses on both the demand side and the supply side are confirmed by the regression results. On the demand side, coefficients on the objective flow are positive and significant in all models, even after controlling for performance and objective size. A one standard deviation increase in the rank of the objective flow increases the likelihood of a new ETF launch by 2.97%–4.93% across the three model specifications. In the meantime, coefficients on the performance measures, i.e. the benchmark-adjusted returns of the family and the investment objective, rarely show any significance. This suggests that families pay relatively little attention to prior performance in making ETF launching decisions, instead they rely more on the prior flows to gauge the growth potential of an objective market. This finding is in direct

[12] The two measures of size and scope are highly colinear with a pairwise correlation of $\rho = 0.81$, as a result the two variables shall not be included in the same model.

contrast to that of open-ended MFs (Khorana and Servaes, 1999). Next, I explore liquidity features that are unique to ETFs as they are traded on the stock exchanges. In all four models, the objective dollar volume appears to be a significant determinant on the ETF launching decisions both statistically and economically. Dollar volume on the family level also appears to be positively correlated with the family decision to launch a new ETF. The economic significance is particularly evident in Model (2), where a one standard deviation increase in the objective and family dollar volume leads to a 8.57% and 11.99% increase in the annualized probability of new ETF launches, respectively. These results suggest that families and objectives who are able to attract more investor attention and hence enjoy higher liquidity are more likely to witness new entrants of ETFs.

On the supply side, the positive and significant relation between the objective tracking error and the likelihood of launching a new ETF is persistent across all four model specifications. The probability of a family launching a new ETF in a certain objective increases by 1.87%-5.18% annually with a one standard deviation increase in the average tracking error of the investment objective. This empirical result confirms our conjecture that fund families may make indirect compensation to the APs to attract market makers by launching new ETFs in objectives with more arbitrage opportunities.

Model (2) and (3) show that the average family expense ratio has a positive and significant impact on the ETF launching decision of fund families. A one standard deviation increase in the family expense ratio increases the annualized probability of a new ETF launch by 5.46%-7.83%. The average expense ratio of the investment objective also affects the likelihood of ETF launching decisions positively though not as significant. The above shows that families are indeed concerned with the ability to generate additional fee income when making ETF launching decisions, which is consistent with the open-ended MF literature (Khorana and Servaes, 1999).

Next, I investigate whether families care about cost reductions on top of revenue growth. Cost reduction can be achieved by a fund family through both cost-sharing on the fixed expenditures and savings on the variable trading costs. Cost-sharing can be realized through economies of scale and scope, which I discuss in more details in the next section. Meanwhile, the activeness of a fund can be measure by the portfolio turnover, a lower portfolio turnover indicates less trading activities and hence lower transaction cost. Coefficients from Model (2)-(4) confirm a negative and significant relationship between the likelihood of new ETF launches and the average portfolio turnover on both the family level and the investment objective level. The annualized probability of new ETF launches decreases by 2.83%-5.99% with a one standard deviation increase in objective portfolio turnover, and 1.77%-6.17% with a one standard deviation increase in family portfolio turnover across the three model specifications.

In addition to factors pertaining to the ETF market, I also investigate whether family decisions to launch new ETFs are affected by characteristics of the corresponding objective in the open-ended MF market. Families may pay attention to prior flows into the MF objective as it signals the investor sentiments towards the asset class. ETFs generally charge lower fees than open-ended MFs, while the average expense ratios of ETFs and open-ended MFs within the same investment objectives are positively correlated, with a pairwise correlation of $\rho = 0.58$. Given the competition on investor flows between the active and passive world, the size of the MF objective measures the potential size of the capital pool. Coefficient estimates in Model (4) show that objective-level flow, expense ratio and size measures of open-ended MFs all appear to have positive and significant impact on the likelihood of new ETF launches. Comparing with the results in Model (2), I find that these objective-level measures of open-ended MFs affect ETF launching decisions in the same direction as measures of the corresponding ETF objectives, though with smaller economic magnitude. Poor risk-adjusted performance of the active funds may make the investors doubt if the active management add value and hence drive investors to the passive alternatives. As a measure of the objective level performance of the open-edned MFs, I compute the Carhart four-factor alpha using a prior 36-month rolling window and take the prior 12-month average of all funds within the same investment objective. However, the empirical evidence does not support this intuition, no statistical significance is shown from the regression coefficient.

Overall, the results in Table 2.2 confirm both Hypothesis 1 and 2. On the demand side, investor demand is regarded more seriously by ETF providers than past performance. Families pay more attention to flows and trading volumes in making ETF launching decisions. On the supply side, families are willing to launch in objectives with more arbitrage opportunities to attract more participation from the market makers.

2.4.2 Economies of Scale and Scope

I find strong support to the economies of scale and scope argument which is persistent across all model specifications in both the full sample and the subsample of existing families. There are significant cost complementarities across ETF products in different objectives within a family. Families are able to effectively reduce the per unit cost through cost sharing on research, operation, marketing and distribution, etc. Also, families who have gone through the launching process in the previous period could save on the fixed costs associated with product development. In the whole sample, Model (2) and (3) from Table 2.2 show that a one standard deviation increase in the family size and the number of objectives covered in the family increases the annualized probability of new ETF launches by 15.49% and 5.90% respectively.

Recognizing that the decision to launch an ETF by a new family might differ from that of an existing family, in Table 2.3 I look at the subsample of only the existing families. The positive and significant impact of family size and scope on ETF launching decisions persists. In addition, an ETF launched in the prior year by a family increases the probability of a new ETF launch by 9.65%-16.35% across the two model specifications, showing that families benefit from prior launching experience. In Model (1) of Table 2.3, the binary variable indicating a new objective without prior family investment demonstrates the strongest impact on the ETF launching decision, which is consistent with the idea of expanding the breadth of fund offerings to exploit the economies of scope. The probability of family making new launches in an objective increases by 25.60% annually if the new fund offering broadens the product line of the family. Next, I explore families' desire to benefit from specialization when launching a new ETF within their existing investment objectives. To approach this problem, I first look at the objective distribution within existing families. Figure 2.3 presents the number of ETF offerings in each objective against the percentage family asset invested for each of the Big Three providers.[13] The positive relation between the number of funds launched into a given investment objective and the percentage family TNA invested indicates the desire by families to specialize in objectives of their expertise and capitalize on the economies of scale. Model (2) in Table 2.3 confirms that families are more likely to launch an additional fund into an investment objective when the objective already represents a large percentage of the family assets. This pattern is consistent with the findings by Evans (2010) but inconsistent with the findings by Khorana and Servaes (1999) for equity MFs.

The presence of significant scale and scope economies may benefit the larger families, and at the same time impose pressure on new market entrants to expand quickly and gain a competitive edge. I study how families balance between the need to expand the breadth of their offerings and the desire to benefit from specialization and the associated economies of scale. Under a multinomial logistic regression framework, I examine the determinants of family decisions to launch an ETF in a new objective versus existing objective in the family, both relative to the decision of no launching. The dependent variable takes a value of one if in a given month the family launches an ETF in a new objective with no prior investment within the family. It takes a value of two if the family launches in an existing objective and zero otherwise. All explanatory variables are standardized to have a mean of zero and standard deviation of one. Calendar year fixed effects are included, and the standard errors are clustered by fund family. Under each model specification in Table 2.4, the first column presents the coefficient estimates corresponding to the determinants of launching in a new objective. The second column presents the results corresponding to the determinants of launching in an existing objective within the family. The third column presents the p-value of a difference test between the coefficients in the first two columns, which sheds light on how the family and objective-level characteristics affect the decisions to launch in new and existing objectives differently.

Column (1) and (3) in Table 2.4 suggest that the decision by existing families to launch in

[13]For the labels of the investment objectives to be visible to readers, only the top ten objectives according to the percentage family asset invested are reported in the graphs. Notice that there is a particular investment objective according to Morningstar classifications, namely S&P 500 Tracking, which represents a substantial (if not the largest) proportion of family TNA in all three largest index providers. Yet only one fund is needed to track a single index in each family, hence the S&P 500 Tracking objective is removed from the figures.

a new objective without previous family investment is dominated by smaller-sized families and families with smaller scope, while the statistical significance on all other explanatory variables largely disappeared.[14] Column (2) and (4) in Table 2.4 suggest that factors affecting the decision of existing families to launch in existing objectives within the family largely conform with those of the full sample. A one standard deviation increase in the family size is associated with a 8.72% decrease in the annualized probability of family launching in a new objective, versus a 16.11% increase in the annualized probability of family launching in existing objectives, both relative to no launching.[15] The p-values from the difference tests show that families who possess larger scale and scope, who enjoy higher average dollar volume and who are able to charge higher expense ratios are more likely to launch in existing objectives. The family decision to launch within an objective is more sensitive to the size of the objective, flows into existing fund offerings in the same objective and average dollar volume of ETFs within the same objective if the family already have ETF offerings in that objective than if it represents a new objective without previous family investment. Overall, the results from Table 2.3 and 2.4 are consistent with the notion in Hypothesis 3 that families with smaller scope face the pressure to expand the breadth of offerings to withstand market competition which favours the larger families, while more established families benefit from specialization and the associated economies of scale.

2.4.3 Role of the Big Three in the ETF Market

In this section, I examine the impact of the Big Three on competition within the ETF market and test Hypothesis 4 on the behaviour of non-Big Three families. In Table 2.5, I repeat the panel logistic regression of Table 2.2 in the subsample of non-Big Three families.

[14]Family size and the number of objectives in the family are included separately in two different model specifications due to the collinearity concern mentioned before (pairwise correlation of the two variables is $\rho = 0.81$).

[15]The marginal effect of family size is calculated by setting all other explanatory variables at the mean (the mean equals zero as the variables are standardized) and indicator variables at zero. For instance,

$$\left(\frac{e^{-0.827-(-4.331)}}{1+e^{-0.827-(-4.331)}} - \frac{e^{-4.331}}{1+e^{-4.331}}\right) * 12 * 100\% = -8.72\%$$

.

All the family and objective-level characteristics specified in Table 2.2 are included as controls. The main variable of interest is the binary variable indicating the Big Three launched in a certain objective during the prior year. The positive and significant coefficients under both model specifications in Table 2.5 show that families tend to follow the leader in picking asset classes for their new launches. An ETF launched by the Big Three in the prior year increases the probability that non-Big Three families launch a new ETF in the same objective by 4.83% and 3.96% annually according to the two model specifications. The statistical pattern on most control variables persists though the economic significance is dampened.

To the extent that product innovation in the ETF industry could be imitated without incurring significant costs and the follower could benefit from the overall growth of the market, it is rational for the non-Big Three families to follow the leader and take the second mover advantage (Reinganum, 1985). However, the profitability of the second mover is related to factors such as brand loyalty and switching cost in the market. To examine whether families are concerned with such frictions when pursuing follow-the-leader strategy, I interact the indicator variable for Big Three launched in the pior year with the Big Three market share and bid-ask spread in the given objective. The negative and significant coefficient on the interaction term in Model (1) shows that non-Big Three families are less likely to launch an ETF into objectives where the market is dominated by the Big Three, even if the Big Three launched in the prior year. This is consistent with the notion that the presence of significant scale economies favours the Big Three. It is easier for the Big Three to develop brand recognition and customer loyalty especially in markets where they possess the monopoly power. Hence the first mover advantage of the Big Three is more likely to sustain in such objectives. Result in Model (2) shows that non-Big Three families are less likely to launch in an investment objective where the bid-ask spread is high, even if the Big Three launched in the prior year. Higher bid-ask spread signals less liquidity in the investment objective and hence higher switching cost for the existing customers, which makes the second movers harder to compete. Overall, empirical evidence from Table 2.5 supports the conjecture in Hypothesis 4.

2.4.4 Lifespan of the ETF Offerings

In this last section of analysis, I use a time-to-event analysis to examine the survival time of ETF offerings. More specifically, I am interested in whether family and objective-level characteristics upon inception could predict the lifespan of the newly-launched ETFs. The lifetime of ETFs is measured in months and right-censored. In our full sample, the lifetime of ETFs has a distribution of mean 43 and median 30, which is equivalent to an average life of 3.6 years and median life of 2.5 years. In Table 2.6, I apply a fully parametric Accelerated Failure Time model with Weibull distribution to allow for a nonlinear hazard function, noticing that the hazard rate of ETF offerings may be higher in the earlier stage and decreases at later times. As before, all explanatory variables are standardized to have a mean of zero and standard deviation of one.

In Model (1) of Table 2.6, I look at the family-level characteristics. Results show that a one standard deviation increase in the logarithm of the family size and family expense ratio extends the lifespan of an ETF offering by 35.53% and 709.30% respectively.[16] In the meantime, a one standard deviation increase in the flow rank of the family shortens the survival time of the ETF offering by 95.16%. Similarly, on the objective level, a one standard deviation increase in the dollar volume and expense ratio shortens the survival time of the ETF offering by 65.87% and 68.21% respectively in Model (2). Both the statistical and economic significance of the aforementioned characteristics persist in Model (3), where both the family and objective-level characteristics are included simultaneously.

Overall, results in Table 2.6 confirms Hypothesis 5. ETFs launched by larger and higher-fee families, and whose initiations are not driven by capital flows into the families are more likely to survive for longer. ETFs launched in lower-fee objectives and whose asset allocation decision is not affected by large trading activities on the market are more likely to live a longer life.

[16] The marginal effect of family size on the survival time in Model (1) is calculated as $e^{0.304} - 1 = 35.53\%$. All other coefficients are interpreted in the same way.

2.5 Robustness Check

In unreported results, I check different measures of performance. The benchmark-adjusted returns are replaced by the excess returns and the value-added measure proposed by Berk and Green (2004).[17] None of the performance measures shows any persistent statistical significance towards the probability of new ETF launches. Also, I examine both the equally-weighted average and the asset-weighted average on returns, expense ratios and portfolio turnovers, etc. The statistical patterns are largely unaffected by the weighting method. For the sake of brevity, only the equally-weighted measure is reported in the tables.

In order to focus on the comparison between active and passive investment decisions, all actively managed ETFs and indexed MFs are excluded from the analysis in Table 2.2. However, given that there are 152 actively managed ETFs in the sample, contributing to around 10% of the total AUM, it is also worth checking whether the determinants on active ETF launches are significantly different from those passive ETFs. In Table 2.7, I perform a multinomial logistic regression to compare the launching decisions of active versus passive ETFs, both relative to no launching. The dependent variable takes a value of one if an active ETF is launched by a family in the given objective during a given month. It takes a value of two if a passive ETF is launched, and zero otherwise. Calendar year fixed effects are included, and the standard errors are clustered by fund family. The p-values from the difference tests suggest little difference between the two model specifications on major characteristics, except that active ETFs are more likely to be launched in smaller objectives.

Recognizing the triopoly structure of the ETF market, I examine whether the factors affecting ETF launching decisions are different when it comes to the Big Three. In Table 2.8, I interact the Big Three dummy with all major characteristics on the family and objective-level. Coefficients on the interaction terms are in general positive and insignificant, suggesting that determinants on the Big Three launching decisions are similar to the whole sample.

Funds whose listing date is at least 12 months after the inception date are classified as incubated, following Evans (2010). In the full sample, there exist 124 such cases with complete

[17] Formula for the value-added: $Value\ Added = TNA * (r_{GROSS} - r_{BMK})$.

data, representing less than 10% of the total number of funds. Evans (2010) argue that MF families apply the incubation strategy to increase performance and attract flows. Intuitively, incubation should be less of a concern for ETFs. Investors would not rely on the track record of a passive provider as much as an active one, as long as they believe the fund would closely track the claimed index.[18] However as a robustness check, I examine whether certain characteristics affect family decisions to incubate when making new ETF launches. Under a multinomial logistic regression framework, I examine the determinants of family decisions to launch an incubated ETF versus a non-incubated ETF, both relative to the decision of no launching. The dependent variable takes a value of one if a family launches an incubated fund in the given objective during a given month. It takes a value of two if the family launches a non-incubated fund and zero otherwise. As before, calendar year fixed effects are included, and the standard errors are clustered by fund family. Results from this multinomial logistic regression are shown in Table 2.9, with the first column showing determinants of an incubated fund, second column showing determinants of a non-incubated fund, and the last column showing the p-value of the difference test on the coefficients in the first two columns. Results suggest there are hardly any differences between the incubated and non-incubated launches. Hence, the main findings are robust to family decisions on incubation.

2.6 Conclusion

In this paper, I investigate the factors that determine family decisions to launch new ETFs and compare with the determinants on the launching decisions of open-ended MFs. ETF launching decisions by 155 families across 74 investment objectives over a 23-year period from January 1996 to December 2018 are examined. The empirical patterns are persistent and robust to different model specifications and subsample tests, which lend strong support to the following conclusions.

[18]Todd Rosenbluth, head of ETF and mutual fund research at CFRA, said that "People believe they understand what they are getting with an ETF that is index based and transparent. Investors are responding quickly to new products and not treating them like a mutual fund or a fine wine needing to age well". The Financial Times article could be found here: https://www.ft.com/content/a611821b-df9b-46d3-acca-fcce6d64f601

Families are concerned with profit maximization in their launching decisions. They care about both revenue generation and cost reduction. Revenue growth is achieved through several channels. First, families attract flows and expand their asset base by matching investor demands, that is to launch in objectives enjoying higher flows and trading volumes. Second, families attract market makers (APs) by launching in objectives with higher tracking error and hence more arbitrage profits. Third, families charge higher percentage fees and extract more value from the fee income. In contrast to open-ended MFs, I find no significant impact of past performance on family launching decisions of ETFs. Cost reduction is achieved through two major channels. First, families applying passive strategies with lower portfolio turnover could save on trading costs. Second, families could effectively reduce the per unit cost by exploiting the economies of scale and scope, which are both significant in the ETF industry.

This leads to the second consideration of fund families in making ETF launching decisions. On the one hand, the presence of significant scale economies makes the specialization strategy attractive. On the other hand, the presence of significant scope economies forces new market entrants to expand quickly in order to gain a competitive edge. As a trade-off, smaller families facing the pressure to expand the breadths of their offerings are more likely to launch in a new objective without previous family investment, while larger families are more likely to specialize in objectives of their expertise and benefit from the associated economies of scale.

Similar to open-ended MFs, ETF families also tend to follow the leader, they are more likely to make new ETF launches in objectives where the Big Three has launched in the prior year. However, after a closer look at the market condition of the given objectives, I find that the willingness of Non-Big Three families to follow the leader is reduced significantly when the objective market is less liquid with higher switching cost or when the investment objective is dominated by the Big Three.

Finally, family and objective characteristics upon inception have significant predicting power on the lifespan of ETF offerings. ETFs launched by larger and higher-fee families; whose inception is not driven by excessive flows into the family are more likely to survive for longer. Also, ETFs launched in lower-fee objectives and whose launching decision is not affected by large trading activities in the objective market are more likely to live a longer life.

In summary, this paper helps regulators, asset managers and other market participants understand what determines the launches of ETFs and how the determinants compare to open-ended MFs.

Figure 2.1. Aggregate AUM growth of US equity ETFs and the number of launches versus closures from 1996 to 2018

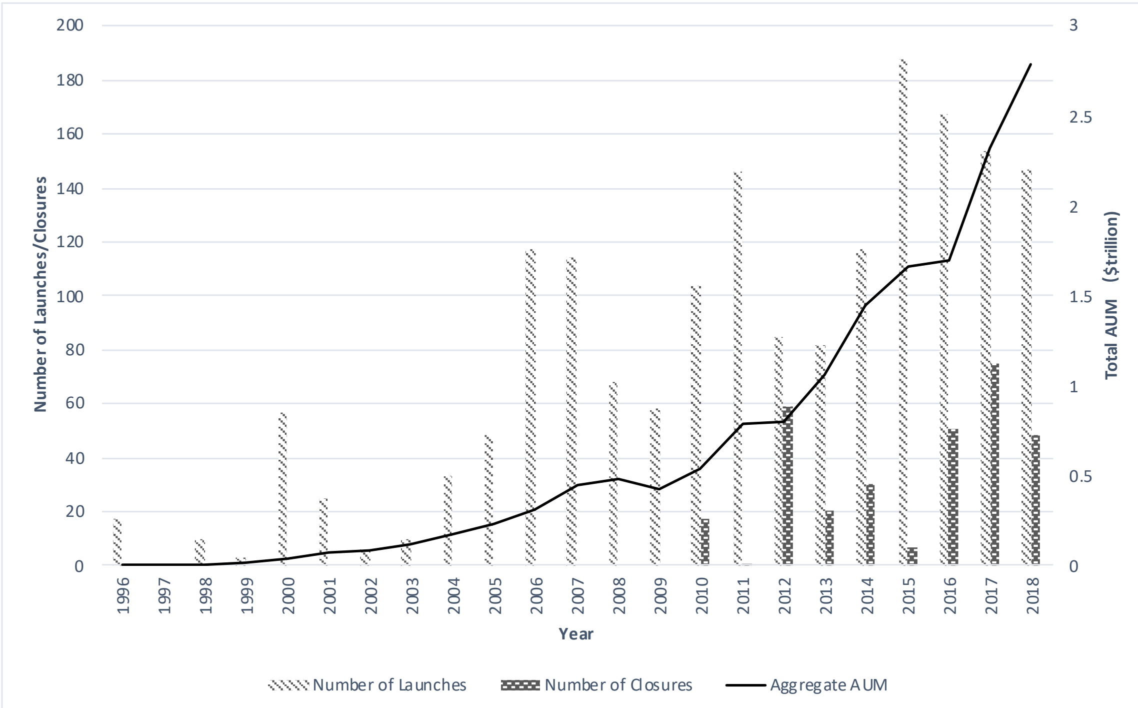


Figure 2.2. Time series of the Big Three market share and the number of launches within and outside the Big Three from 1996 to 2018

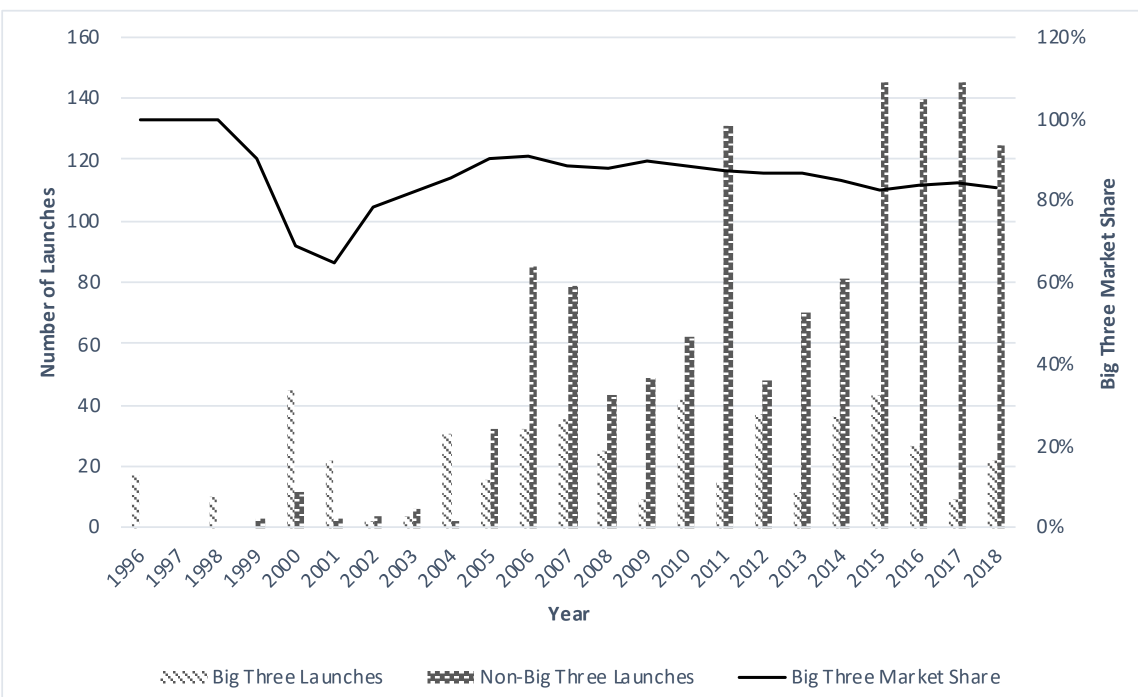


Figure 2.3. Distribution of the number of ETF offerings against the percentage family TNA invested in the top ten investment objectives of the Big Three, as of December 2018

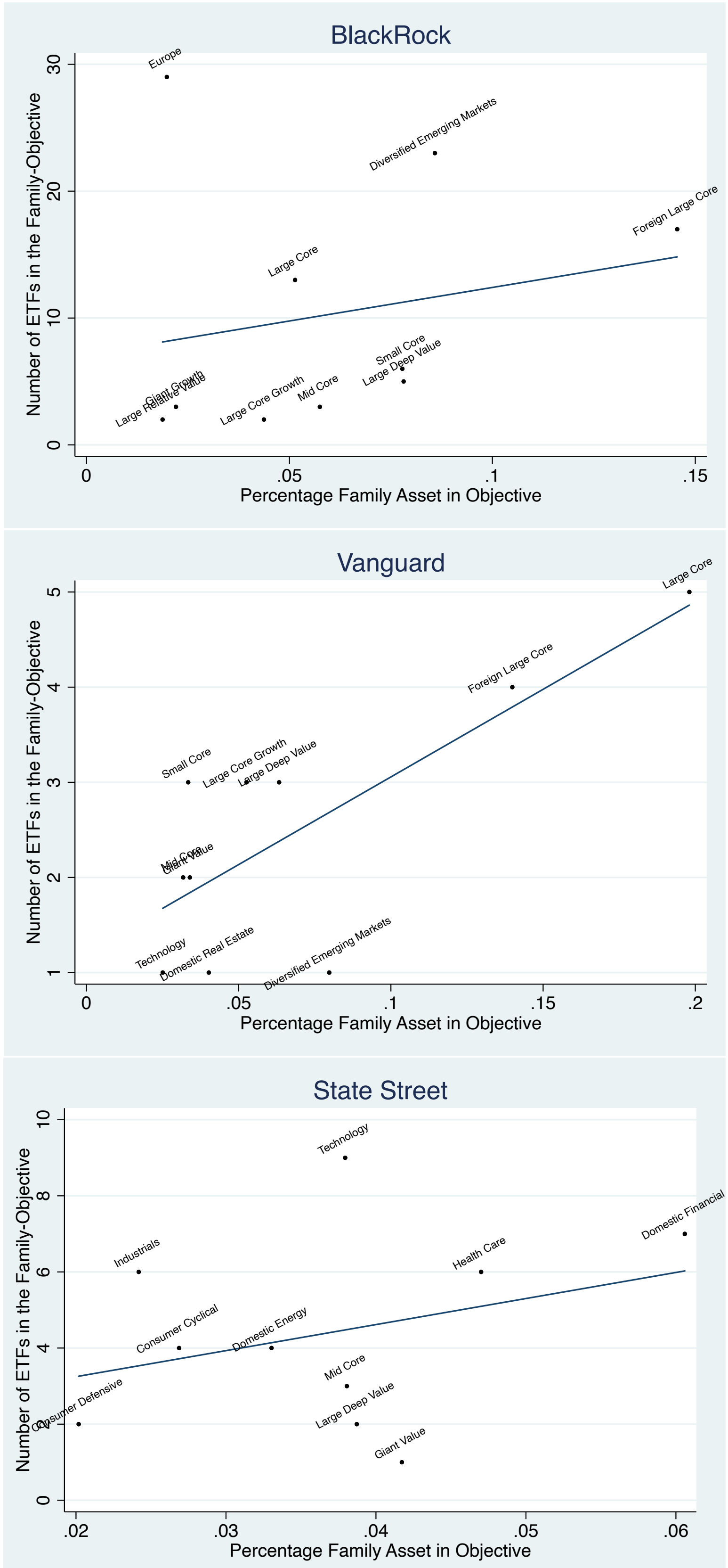


Table 2.1
Descriptive Statistics

This table contains descriptive statistics of the fund-month observations on major fund characteristics. Expense Ratio and Portfolio Turnover are annualized. Dollar Volume and Bid-ask Spread are computed as the monthly average of daily measures and winsorized at the 1% and 99% level. Net Flow is calculated monthly as the percentage change in TNA net of any return effect. Benchmark-Adjusted Return is calculated by regressing the monthly net returns on the primary benchmark returns through a prior 36-month rolling window. Default benchmark of the investment objective (Table A2.2) is used when the primary benchmark is missing. In panel A, summary statistics are reported for all ETFs in the sample. Panel B compares the above characteristics between sub-samples of the Big Three versus non-Big Three families. Panel C compares between the active versus passive sub-samples. The"MeanDiff" column reports the difference in the sample means, together with the statistical significance from a two-sided t-test. The asterisks denote statistical significance as follows: *** significant at 0.1%, ** significant at 1%, and * significant at 5%.

Panel A: Funds from All Families

Variables	Obs.	Median	Mean	Std. Dev.	Min.	Max.
Net Flow ($ millions)	122,421	-0.005	13.681	329.659	-27,705	22,707
Dollar Volume ($ thousands)	41,730	12.592	15.741	11.891	0.677	65.543
Tracking Error (%)	123,998	1.564	2.147	2.096	0.000	25.618
Expense Ratio (%)	122,736	0.480	0.477	0.249	-0.140	5.070
Portfolio Turnover (%)	124,224	22.000	36.002	40.684	0.000	228.000
Fund TNA ($ millions)	126,695	107	1,495	7,472	0.477	306,671
Family TNA ($ millions)	285,288	24,186	123,532	213,839	0.558	1,117,694
Bid-Ask Spread (%)	61,190	0.079	0.133	0.179	0.010	1.261
Benchmark-Adjusted Return (%)	114,310	0.039	0.007	2.956	-187.829	35.933

Panel B: Comparision Between the Big Three and the Non-Big Three

Variables	Non-Big Three		Big Three		MeanDiff
	Obs.	Mean1	Obs.	Mean2	
Net Flow ($ millions)	66,838	4.880	55,583	24.265	-19.385***
Dollar Volume ($ thousands)	29,182	13.803	12,548	20.248	-6.444***
Tracking Error (%)	67,716	2.405	56,282	1.838	0.567***
Expense Ratio (%)	67,116	0.569	55,620	0.366	0.203***
Portfolio Turnover (%)	67,296	49.867	56,928	19.612	30.255***
Fund TNA ($ millions)	69,921	407	56,774	2,834	-2.4e+03***
Family TNA ($ millions)	150,627	21,000	134,661	240,000	-2.2e+05***
Bid-Ask Spread (%)	38,880	0.138	22,310	0.124	0.014***
Benchmark-Adjusted Return (%)	61,933	-0.038	52,377	0.061	-0.099***

Panel C: Comparision Between Active and Passive ETFs

Variables	Passive		Active		MeanDiff
	Obs.	Mean1	Obs.	Mean2	
Net Flow ($ millions)	118,089	14.117	4,332	1.796	12.321**
Dollar Volume ($ thousands)	38,870	15.780	2,860	15.213	0.567**
Tracking Error (%)	119,645	2.166	4,353	1.636	0.530***
Expense Ratio (%)	118,320	0.466	4,416	0.760	-0.293***
Portfolio Turnover (%)	119,844	34.608	4,380	74.166	-39.559***
Fund TNA ($ millions)	122,202	1,547	4,493	69	1477.858***
Family TNA ($ millions)	269,924	130,000	15,364	95,000	3.0e+04***
Bid-Ask Spread (%)	57,791	0.131	3,399	0.169	-0.038***
Benchmark-Adjusted Return (%)	110,297	0.009	4,013	-0.037	0.046

Table 2.2

Panel Logistic Regressions on the likelihood of new ETF Launches

This table presents the results from panel logistic regression models on the likelihood of new ETF launches. The dependent variable is an indicator variable taking a value of one if a family launches in a given investment objective during a given month, and zero otherwise. Various family and objective characteristics of ETFs are included as explanatory variables across Model (1) to (3). In Model (4), objective characteristics are substituted by the MF counterparty. The models assume independence of ETF launching decisions across families, but not within families. Calendar year fixed effects are included, and the standard errors are clustered by fund family. All the explanatory variables are standardized to have a mean of zero and a standard deviation of one. The marginal effect is reported in the column"ME", which is calculated as the annualized percentage change in the probability of a new ETF launch when each explanatory variable is increased by one standard deviation and all other variables are set equal to the mean. For indicator variables, this represents the percentage change in probability when the indicator variable increases from zero to one. ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively.

	All Families							
Variables	Model (1)		Model (2)		Model (3)		Model (4)	
	Coefficients	ME	Coefficients	ME	Coefficients	ME	Coefficients	ME
Objective Flow	0.319***	2.972	0.129***	4.934	0.129***	4.226		
	(0.046)		(0.036)		(0.035)			
Family Flow	0.061	0.499	-0.022	-0.784	0.006	0.185	-0.023	-0.240
	(0.073)		(0.085)		(0.090)		(0.088)	
Objective Dollar Volume	0.098***	0.816	0.215***	8.569	0.212***	7.229	0.122***	1.368
	(0.035)		(0.037)		(0.037)		(0.035)	
Family Dollar Volume	0.120	1.010	0.288***	11.889	0.144**	4.752	0.263***	3.166
	(0.098)		(0.077)		(0.072)		(0.075)	
Objective Tracking Error	0.212***	1.870	0.135***	5.179	0.140***	4.611	0.251***	3.003
	(0.040)		(0.041)		(0.042)		(0.046)	
Objective Benchmark-Adjusted Return	0.062	0.507	0.046	1.692	0.048	1.513		
	(0.054)		(0.054)		(0.056)			
Family Benchmark-Adjusted Return	0.027	0.217	-0.052	-1.826	0.002	0.062		
	(0.070)		(0.071)		(0.070)			
Objective Expense Ratio			0.129*	4.934	0.123*	4.018		
			(0.067)		(0.067)			
Family Expense Ratio			0.198**	7.827	0.164**	5.464	0.189**	2.191
			(0.092)		(0.083)		(0.096)	
Objective Portfolio Turnover			-0.181***	-5.989	-0.177***	-5.018	-0.311***	-2.827
			(0.068)		(0.067)		(0.058)	
Family Portfolio Turnover			-0.187**	-6.171	-0.208***	-5.814	-0.183**	-1.767
			(0.072)		(0.077)		(0.073)	
Objective Size			0.693***	34.898	0.665***	28.387		
			(0.101)		(0.105)			
Family Size			0.362***	15.489			0.276***	3.344
			(0.110)				(0.100)	
Number of Objectives in Family					0.176**	5.898		
					(0.074)			
MF Objective Flow							0.079*	0.867
							(0.043)	
MF Objective Expense Ratio							0.208***	2.435
							(0.060)	
MF Objective Size							0.456***	6.065
							(0.044)	
MF Objective Alpha_FF4F							0.003	0.032
							(0.063)	
Constant	-5.006***		-3.444***		-3.609***		-4.716***	
	(0.189)		(0.275)		(0.345)		(0.187)	
Observations	68,518		67,195		68,228		67,258	
Number of Family-Objectives	767		760		767		758	
Year FE	YES		YES		YES		YES	
Wald χ^2	8.43E+04		1.36E+08		9.98E+07		4.93E+05	

Table 2.3

Economies of Scale and Scope, Benefit of Specialization

This table presents the results from panel logistic regression models on the likelihood of new ETF launches in the subsample of existing families. The dependent variable is an indicator variable taking a value of one if an existing family launches in a given investment objective during a given month, and zero otherwise. In Model (1), *No Prior Family Investment in Objective* is a dummy variable indicating a new objective within the family. In Model (2), only observations within existing families and existing objectives are considered. All the explanatory variables are standardized to have a mean of zero and a standard deviation of one. The marginal effect is reported in the column"ME", which is calculated as the annualized percentage change in the probability of a new ETF launch when each explanatory variable is increased by one standard deviation and all other variables are set equal to the mean. For indicator variables, this represents the percentage change in probability when the indicator variable increases from zero to one. The models assume independence of ETF launching decisions across families, but not within families. Calendar year fixed effects are included, and the standard errors are clustered by fund family. ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively.

Variables	Model (1)		Model (2)	
	Existing Families		Existing Family-Objectives	
	Coefficients	ME	Coefficients	ME
No Prior Family Investment in Objective	1.108***	25.601		
	(0.192)			
Percentage Family Asset in Objective			0.189***	2.358
			(0.048)	
Family Launched in Prior Year	0.562***	9.650	0.900***	16.349
	(0.184)		(0.339)	
Objective Size	0.702***	12.986	1.112***	22.731
	(0.126)		(0.091)	
Family Size	0.367***	5.692	0.863***	15.362
	(0.130)		(0.115)	
Objective Flow	0.091**	1.228	0.143***	1.743
	(0.036)		(0.038)	
Objective Dollar Volume	0.217***	3.118	0.349***	4.725
	(0.047)		(0.042)	
Family Dollar Volume	0.200***	2.849	0.323***	4.314
	(0.077)		(0.097)	
Objective Tracking Error	0.145***	2.009	0.144***	1.756
	(0.055)		(0.056)	
Objective Expense Ratio	0.183**	2.585	0.201**	2.523
	(0.072)		(0.082)	
Family Expense Ratio	0.115	1.570	0.429**	6.053
	(0.113)		(0.190)	
Objective Portfolio Turnover	-0.058	-0.727	-0.218**	-2.229
	(0.062)		(0.107)	
Family Portfolio Turnover	-0.120	-1.461	-0.231***	-2.347
	(0.087)		(0.076)	
Constant	-4.511***		-4.641***	
	(0.349)		(0.559)	
Observations	67,540		48,061	
Number of Family-Objectives	704		669	
Year FE	YES		YES	
Wald χ^2	2.540E+07		2.090E+07	

Table 2.4
Multinomial Logit Regression on the Choice of Objectives

This table presents the results from multinomial logistic regression models on the decision to launch in a new objective versus an existing objective within the family, both relative to no launching. The dependent variable takes a value of one if, in a given month, the family launches a new ETF in a new objective with no prior investment within the family. It takes a value of two if the family launches a new ETF in an existing objective of the family in that given month and takes a value of zero otherwise. The column "Difference p-value" presents the p-value of a difference test between the coefficients across two model specifications. The models assume independence of ETF launching decisions across families, but not within families. Calendar year fixed effects are included, and the standard errors are clustered by fund family. All the explanatory variables are standardized to have a mean of zero and a standard deviation of one. Pseudo R^2 is computed as one minus the log-likelihood ratio at convergence over the log-likelihood ratio at zero. ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively.

	Existing Families Launch in					
Variables	New Objective	Existing Objective	Difference p-value	New Objective	Existing Objective	Difference p-value
	(1)	(2)	(1) - (2)	(3)	(4)	(3) - (4)
Family Size	-0.827***	1.112***	0.000			
	(0.158)	(0.110)				
Number of Objectives in Family				-1.040***	0.599***	0.000
				(0.115)	(0.078)	
Objective Size	-0.094	1.318***	0.000	-0.162*	1.320***	0.000
	(0.091)	(0.093)		(0.092)	(0.093)	
Objective Dollar Volume	-0.040	0.385***	0.000	-0.064	0.387***	0.000
	(0.054)	(0.048)		(0.050)	(0.047)	
Family Dollar Volume	-0.118	0.359***	0.001	0.109	0.014	0.411
	(0.124)	(0.092)		(0.089)	(0.079)	
Objective Tracking Error	-0.101	0.158***	0.014	-0.085	0.162***	0.034
	(0.087)	(0.047)		(0.088)	(0.052)	
Objective Flow	0.104*	0.121***	0.801	0.083	0.120***	0.576
	(0.061)	(0.037)		(0.061)	(0.035)	
Objective Expense Ratio	0.120	0.263***	0.226	0.098	0.267***	0.173
	(0.095)	(0.082)		(0.097)	(0.083)	
Family Expense Ratio	0.040	0.420**	0.026	-0.025	0.299**	0.026
	(0.074)	(0.172)		(0.075)	(0.146)	
Objective Portfolio Turnover	-0.052	-0.172	0.403	-0.044	-0.181	0.328
	(0.054)	(0.126)		(0.052)	(0.125)	
Family Portfolio Turnover	-0.223*	-0.142**	0.512	-0.149	-0.174**	0.803
	(0.114)	(0.065)		(0.098)	(0.069)	
Family Launched in Prior Year	0.388*	0.967***	0.067	0.521***	1.157***	0.085
	(0.204)	(0.333)		(0.175)	(0.373)	
Constant	-4.331***	-4.997***		-4.135***	-5.770***	
	(1.029)	(0.397)		(0.508)	(0.458)	
Observations		67,540			68,427	
Year FE		YES			YES	
Pseudo R^2		0.094			0.099	

Table 2.5

Impact of the Big Three on Competition

This table presents the results from panel logistic regression models on the likelihood of new ETF launches in the subsample of non-Big Three families. The dependent variable is an indicator variable taking a value of one if a family launches in a given investment objective during a given month, and zero otherwise. *Bigthree Launched in Prior Year* is a dummy variable taking a value of one if an ETF was launched by the Big Three in a given objective in the prior year, and zero otherwise. Interaction terms are expressed with "#" between the two explanatory variables. The models assume independence of ETF launching decisions across families, but not within families. Calendar year fixed effects are included, and the standard errors are clustered by fund family. All the explanatory variables are standardized to have a mean of zero and a standard deviation of one. ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively.

Variables	Non-Big Three	
	Model (1)	Model (2)
Bigthree Launched in Prior Year	0.993***	0.942***
	(0.210)	(0.270)
Bigthree Launched in Prior Year # Bigthree Objective Market Share	-0.736***	
	(0.139)	
Bigthree Launched in Prior Year # Objective Bid-Ask Spread		-1.025***
		(0.274)
Objective Flow	0.116**	0.120
	(0.053)	(0.076)
Objective Dollar Volume	0.131**	0.272***
	(0.054)	(0.073)
Family Dollar Volume	0.229***	0.312***
	(0.079)	(0.101)
Objective Tracking Error	0.094	-0.301
	(0.057)	(0.198)
Objective Expense Ratio	0.085	0.136
	(0.083)	(0.187)
Family Expense Ratio	0.048	-0.014
	(0.097)	(0.143)
Objective Portfolio Turnover	-0.049	0.076
	(0.043)	(0.069)
Family Portfolio Turnover	-0.108	-0.072
	(0.070)	(0.097)
Objective Size	0.550***	0.637***
	(0.095)	(0.126)
Family Size	0.237**	0.199
	(0.119)	(0.132)
Family Launched in Prior Year	0.497**	0.420
	(0.229)	(0.263)
Constant	-5.770***	-6.305***
	(0.305)	(0.365)
Observations	44,934	23,252
Number of Family-Objectives	625	598
Year FE	YES	YES
Wald χ^2	7295.2	279.6

Table 2.6
Time-to-Event Analysis on the Lifespan of ETF Offerings

This table reports the results from a time-to-event analysis on the lifespan of the ETF offerings. The dependent variable is the life of an ETF offering, measured in months and right-censored. The specification is an Accelerated Failure Time model with Weibull distribution. All the explanatory variables are standardized to have a mean of zero and a standard deviation of one. ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively.

Variables	Life of ETF		
	Model (1)	Model (2)	Model (3)
Family Flow	-3.033***		-2.611***
	(0.375)		(0.424)
Family Dollar Volume	-1.185***		-0.622
	(0.413)		(0.431)
Family Expense Ratio	2.091***		1.934***
	(0.359)		(0.410)
Family Portfolio Turnover	0.087		0.264
	(0.260)		(0.280)
Family Size	0.304***		0.293***
	(0.038)		(0.039)
Family Launched in Prior Year	-0.257		-0.136
	(0.204)		(0.203)
Family Benchmark-Adjusted Return	0.008*		0.007*
	(0.004)		(0.004)
Objective Flow		0.520**	0.301
		(0.245)	(0.239)
Objective Dollar Volume		-1.075***	-0.452**
		(0.217)	(0.222)
Objective Expense Ratio		-1.146**	-1.624***
		(0.525)	(0.515)
Objective Portfolio Turnover		-1.352***	-0.674*
		(0.281)	(0.345)
Objective Size		-0.135***	-0.035
		(0.040)	(0.042)
Objective Tracking Error		0.076*	0.072*
		(0.042)	(0.041)
Objective Benchmark-Adjusted Return		-0.010	-0.007
		(0.006)	(0.006)
Constant	-0.848	9.876***	0.564
	(0.908)	(0.931)	(1.317)
Observations	1,239	1,468	1,056
Wald χ^2	249.3	63.6	217.0

Table 2.7

Multinomial Logistic Regressions on the Launches of Active versus Passive ETFs

This table presents the results from a multinomial logistic regression model on the decision to launch an active versus passive ETF, both relative to no launching. The dependent variable takes a value of one if a family launches an active ETF in a given investment objective during a given month. It takes a value of two if the family launches a passive ETF and takes a value of zero otherwise. All the explanatory variables are standardized to have a mean of zero and a standard deviation of one. The column "Difference p-value" presents the p-value of a difference test between the coefficients across two model specifications. The models assume independence of ETF launching decisions across families, but not within families. Calendar year fixed effects are included, and the standard errors are clustered by fund family. Pseudo R^2 is computed as one minus the log-likelihood ratio at convergence over the log-likelihood ratio at zero. ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively.

Variables	Active (1)	Passive (2)	Diffefence p-value (1) - (2)
Objective Size	0.144	0.688***	0.010
	(0.237)	(0.115)	
Family Size	0.193	0.378***	0.619
	(0.335)	(0.119)	
Objective Benchmark-Adjusted Return	0.390	0.037	0.311
	(0.331)	(0.055)	
Family Benchmark-Adjusted Return	0.193	-0.041	0.290
	(0.215)	(0.063)	
Objective Flow	0.107	0.124***	0.910
	(0.137)	(0.036)	
Family Flow	0.111	-0.014	0.619
	(0.214)	(0.081)	
Objective Dollar Volume	0.451***	0.216***	0.106
	(0.143)	(0.043)	
Family Dollar Volume	-0.009	0.343***	0.114
	(0.200)	(0.093)	
Objective Tracking Error	-0.476	0.126***	0.099
	(0.375)	(0.037)	
Objective Expense Ratio	-0.140	0.195***	0.302
	(0.326)	(0.063)	
Family Expense Ratio	0.494***	0.098	0.105
	(0.167)	(0.134)	
Objective Portfolio Turnover	0.043	-0.138*	0.312
	(0.153)	(0.074)	
Family Portfolio Turnover	-0.110	-0.228**	0.572
	(0.183)	(0.093)	
Constant	-17.383***	-3.329***	
	(1.762)	(0.226)	
Observations	71,702		
Number of Active/Passive Launches	121 / 1,425		
Year FE	YES		
Pseudo R^2	0.057		

Table 2.8
Robustness Check on Determinants of Big Three Launching Decisions

This table presents the results from panel logistic regression models on the likelihood of new ETF launches in the subsample of existing families. The dependent variable is an indicator variable taking a value of one if an existing family launches in a given investment objective during a given month, and zero otherwise. All explanatory variables in Table 2.2 are included as control variables. The reported coefficients are estimated on the interaction terms of each explanatory variable with the *Big Three* dummy. The models assume independence of ETF launching decisions across families, but not within families. Calendar year fixed effects are included, and the standard errors are clustered by fund family. All the explanatory variables are standardized to have a mean of zero and a standard deviation of one. ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively.

Variables	Model (1)	Model (2)
(Interaction with the Bigthree Dummy)	Existing Families	Existing Family-Objectives
Percentage Family Asset in Objective		0.449**
		(0.191)
Objective Flow	0.015	-0.109*
	(0.049)	(0.064)
Family Flow	-0.167	-0.169
	(0.115)	(0.171)
Objective Dollar Volume	0.244***	0.126
	(0.045)	(0.088)
Family Dollar Volume	-0.002	-0.056
	(0.168)	(0.179)
Objective Tracking Error	0.090	0.097
	(0.123)	(0.151)
Objective Expense Ratio	0.170*	0.140
	(0.093)	(0.135)
Family Expense Ratio	0.749***	0.933***
	(0.184)	(0.319)
Objective Portfolio Turnover	-0.129	-0.143**
	(0.100)	(0.072)
Family Portfolio Turnover	0.865	2.306**
	(0.742)	(0.983)
Objective Size	0.465***	0.200
	(0.181)	(0.199)
Family Size	0.453*	1.082**
	(0.253)	(0.429)
Constant	-4.074***	-5.193***
	(0.428)	(0.717)
Observations	67,540	48,061
Controls Included	YES	YES
Year FE	YES	YES

Table 2.9

Multinomial Logistic Regression on Determinants to Incubate

This table presents the results from multinomial logistic regression models on the decision to launch an incubated ETF versus a non-incubated ETF, both relative to no launching. The dependent variable takes a value of one if, in a given month, the family launches an incubated ETF. It takes a value of two if the family launches a non-incubated ETF and zero otherwise. The column "Difference p-value" presents the p-value of a difference test between the coefficients across two model specifications. The models assume independence of ETF launching decisions across families, but not within families. Calendar year fixed effects are included, and the standard errors are clustered by fund family. All the explanatory variables are standardized to have a mean of zero and a standard deviation of one. Pseudo R^2 is computed as one minus the log-likelihood ratio at convergence over the log-likelihood ratio at zero. ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively.

Variables	Incubated (1)	Non-Incubated (2)	Diffefence p-value (1) - (2)
Objective Flow	0.026	0.138***	0.378
	(0.115)	(0.035)	
Family Flow	0.223	0.007	0.425
	(0.289)	(0.077)	
Objective Dollar Volume	0.161	0.236***	0.714
	(0.195)	(0.044)	
Family Dollar Volume	0.523	0.212***	0.479
	(0.433)	(0.071)	
Objective Tracking Error	-0.019	0.113**	0.325
	(0.121)	(0.055)	
Objective Expense Ratio	0.211*	0.170**	0.758
	(0.112)	(0.070)	
Family Expense Ratio	0.395*	0.116	0.148
	(0.206)	(0.102)	
Objective Portfolio Turnover	-0.554***	-0.074	0.011
	(0.192)	(0.056)	
Family Portfolio Turnover	0.242	-0.202**	0.050
	(0.227)	(0.086)	
Objective Size	0.725***	0.615***	0.508
	(0.097)	(0.138)	
Family Size	0.475	0.234**	0.616
	(0.485)	(0.115)	
Family Launched in Prior Year	1.542***	0.523***	0.073
	(0.521)	(0.196)	
Objective Benchmark-Adjusted Return	0.092	0.026	0.376
	(0.075)	(0.028)	
Family Benchmark-Adjusted Return	0.023	-0.025	0.771
	(0.157)	(0.058)	
Constant	-4.716***	-4.293***	
	(0.814)	(1.029)	
Observations	71,423		
Number of Incubated/Non-Incubated Launches	124 / 1,423		
Year FE	YES		
Pseudo R^2	0.058		

Table A2.1
Hypotheses

This table outlines the ex-ante expectations regarding the effect of each explanatory variable on the likelihood of new ETF launches. The hypotheses are divided into three broad categories, namely profit maximization, scale and scope economies and the impact of the Big Three. Predictions are posed in either the whole sample or two different subsamples of the existing families and the non-Big Three families where appropriate. Interaction terms are expressed with a "#" between the two different explanatory variables.

	Likelihood of new ETF Launches		
Variables	All Families	Existing Families	Non-Big Three Families
Panel A: Profit Maximization			
Revenue Generation			
Flow_Objective (ETF&MF)	Positive		
Flow_Family		Positive	
Dollar Volume_Objective	Positive		
Dollar Volume_Family		Positive	
Tracking Error_Objective	Positive		
Expense Ratio_Objective (ETF&MF)	Positive		
Expense Ratio_Family		Positive	
Performance_Objective	Insignificant		
Performance_Family		Insignificant	
Cost Reduction			
Portfolio Turnover_Objective	Negative		
Portfolio Turnover_Family		Negative	
Panel B: Scale and Scope Economies			
Size_Objective (ETF&MF)	Positive		
Size_Family		Positive	
Scope_Family		Positive	
Prior Launch in Family		Positive	
Benefit of Specialization			
Percentage Family Asset in Objective		Positive	
Panel C: Impact of the Big Three			
Prior Launch in Big Three			Positive
Market Share # Prior Launch in Big Three			Negative
Bid/Ask Spread # Prior Launch in Big Three			Negative

Table A2.2

Default Benchmark of the Investment Objectives

This table lists the default benchmark for each of the 74 investment objectives in our sample. These are obtained by taking the most commonly used (sorted on the TNA) primary benchmark of ETFs in each Morningstar category. I use the default benchmark to compute the benchmark-adjusted returns where the primary benchmark of an ETF is missing.

Investment Objective	Benchmark	Investment Objective	Benchmark
All-Cap Core	S&P 1500 PR	Large Relative Value	MSCI USA Minimum Volatility (USD) NR USD
All-Cap Growth	Russell 3000 Growth NR USD	Large Valuation-Sensitive Growth	MSCI USA Momentum NR USD
All-Cap Value	NASDAQ US Dividend Achievers 50 PR USD	Latin America	MSCI ALL Argentina 25/50 PR USD
China	MSCI China GR USD	Leveraged	Russell 2000 NR USD
Commodities Agriculture	Rogers Intl Commodity Agriculture TR USD	Leveraged Net Long	Russell 1000 Leverage Long PR USD
Communications	MSCI US IMI/Comm Svc 25-50 NR USD	Materials	Morningstar Gbl Upstm Nat Res PR USD
Consumer Cyclical	S&P Consumer Disc Select Sector PR USD	Micro Cap	Russell Micro Cap PR USD
Consumer Defensive	S&P Cons Staples Select Sector PR USD	Mid Core	S&P MidCap 400 NR
Diversified Emerging Markets	MSCI EM GR USD	Mid Core Growth	S&P MidCap 400 Growth TR USD
Diversified Pacific/Asia	FTSE Dvlp Asia Pacific (US RIC) NR USD	Mid Core Value	Russell 2000 Value NR USD
Domestic Energy	Alerian MLP Infrastructure PR USD	Mid Deep Value	S&P MidCap 400 Value TR USD
Domestic Financial	S&P Financial Select Sector TR USD	Mid High Growth	Morningstar US Mid Growth PR USD
Domestic Real Estate	DJ US Select REIT PR USD	Mid Relative Value	S&P MidCap 400 Equal Weighted TR USD
Emerging Europe	MVIS Russia NR USD	Pacific/Asia ex-Japan	MSCI Korea 25-50 100% Hedged NR USD
Emerging Markets Bond	MSCI Pacific GR USD	Precious Metals	NYSE Arca Gold Miners PR USD
Europe	EURO STOXX 50 PR USD	S&P 500 Tracking	S&P 500 NR USD
Foreign Giant	MSCI EAFE Free Value GR USD	SMID Core	S&P Completion TR USD
Foreign Large Core	MSCI EAFE PR USD	SMID Growth	MSCI ACWI SMID Growth GR USD
Foreign Large Growth	MSCI EAFE Free Growth GR USD	SMID Value	MSCI ACWI SMID Value NR USD
Foreign Large Value	DJ EPAC Select Dividend TR USD	Small Core	S&P SmallCap 600 NR USD
Foreign Real Estate	DJ Gbl Ex US Select RESI NR USD	Small Core Growth	S&P SmallCap 600 Growth TR USD
Foreign Small/Mid Core	MSCI EAFE Small Cap 100% Hedge NR USD	Small Core Value	S&P SmallCap 600 Value TR USD
Foreign Small/Mid Growth	MSCI Israel IMI Capped GR USD	Small Deep Value	Russell 2000 Value PR USD
Foreign Small/Mid Value	WisdomTree Intl SmallCap Dividend PR USD	Small High Growth	Morningstar US Small Growth TR USD
Giant Core	S&P 100 TR	Small Relative Value	S&P SmallCap 600 Equal Weighted TR USD
Giant Growth	S&P 500 Growth TR USD	Small Valuation-Sensitive Growth	Russell 2000 High Momentum TR USD
Giant Value	DJ Industrial Average PR USD	Technology	S&P 500 NR USD
Health Care	S&P Health Care Select Sector TR USD	Utilities	S&P Utilities Select Sector PR USD
India	MSCI India PR USD	World All-Cap	MSCI ACWI Low Carbon Target PR USD
Industrials	S&P Industrial Select Sector TR USD	World Energy	S&P Global 1200 Energy Sector PR USD
Infrastructure	S&P Global Infrastructure TR USD	World Financial	MSCI Europe/Financials GR USD
Japan	MSCI Japan PR USD	World Large Core	MSCI ACWI GR USD
Large Core	S&P 500 NR USD	World Large Growth	Gavekal Knowledge Leaders Dev Wld PR USD
Large Core Growth	Russell 1000 Growth PR USD	World Large Value	S&P Global 100 NR USD
Large Core Value	S&P 500 NR USD	World Mid Cap	Solactive Global SuperDividend TR USD
Large Deep Value	Russell 1000 Value PR USD	World Real Estate	FTSE EPRA Nareit Developed TR USD
Large High Growth	S&P 500 Pure Growth PR	World Small Cap	MSCI World Small Cap PR USD

Table A2.3

Variable Definitions

Variable	Definition
	Panel A: Fund-level Variables
Net Flow	$Net\ Flow_{i,t} = TNA_{i,t} - TNA_{i,t-1} * (1 + R_{i,t})$, where $R_{i,t}$ denotes the gross return of fund i during month t.
Dollar Volume	$Dollar\ Volume_{i,t} = \frac{1}{T}\sum_{t=1}^{T} Daily\ Trading\ Price_{i,t} * Daily\ Volume_{i,t}$, where T denotes the number of trading days within month t.
Tracking Error	$Tracking\ Error_{i,t} = \sigma(R_{i,t} - R_{BMK_{i,t}})$, where σ denotes the standard deviation.
Expense Ratio	The annual report net expenseratio represents the percentage of fund TNA used to pay for operating expenses, management fees, administrative fees, and all other asset-based costs incurred.
Portfolio Turnover	$Portfolio\ Turnover_{i,t} = \frac{\min(Purchase_{i,t}, Sale_{i,t})}{TNA_{i,t}}$.
Bid-Ask Spread	Daily average of the amount by which the ask price of an ETF share exceeds the bid price.
Benchmark-Adjusted Return	Difference between the realized return and the expected return predicted from a single-index model using prior-36-month rolling window.
	Panel B: Family/Objective-level Variables
Family/Objective Flow	A rank between zero and one assigned to the total net flows into the family/objective, then take the prior 12-month average.
Family/Objective Dollar Volume	A rank between zero and one assigned to the monthly average dollar volumes in the family/objective, then take the prior-12-month average.
Objective Tracking Error	Prior 12-month average tracking errors for all funds within the investment objective.
Family/Objective Expense Ratio	Prior 12-month average of the annualized expense ratios for all funds within the family/objective.
Family/Objective Portfolio Turnover	Prior 12-month average of the annualized portfolio turnovers for all funds within the family/objective.
Family/Objective Size	Logarithm of the total AUM for all funds within the family/objective by the end of last month.
Number of Objectives in Family	Total number of objectives where the family has ETF offerings.
Family/Objective Benchmark-Adjusted Return	Prior 12-month average Benchmark-Adjusted Returns for all funds within the family/objective.
MF Objective Alpha_FF4F	Four-factor alpha for each open-ended MF is calculated as the abnormal return from the Fama-French four factor model using prior 36-month rolling window, then take the prior 12-month average of all funds within the investment objective.
No Prior Family Investment in Objective	An indicator variable equals one if the family has no ETF offerings in the investment objective.
Percentage Family Asset in Objective	$\frac{Total\ AUM\ of\ all\ ETFs\ offered\ by\ the\ family\ in\ the\ objective}{Family\ Size} * 100\%$
Family Launched in Prior Year	An indicator variable equals one if the family made new ETF launches during the previous year, and zero otherwise.
Bigthree Launched in Prior Year	An indicator variable equals one if BlackRock, Vanguard or State Street made new ETF launches during the previous year, and zero otherwise.
Bigthreee Objective Market Share	Percentage of objective assets under control of BlackRock, Vanguard and State Street by the end of last month.

Chapter 3

The More Things Change, The More They Stay The Same: Why Do Mutual Funds Change Sub-advisors?

Julia Arnold **David Chambers** **Pedro A.C. Saffi** **Xinrui Zheng**

September 28, 2021

Abstract

We study the effects of managerial turnover and competition on U.S. sub-advised mutual funds (MFs), using changes of sub-advisors by 426 funds from January 1995 to December 2016. Sub-advised MFs exhibit return-chasing behaviour when making turnover decisions, but these changes neither improve subsequent fund returns and risk measures, nor increase future flows into the fund. Using sub-advisor turnover to change the degree of competition among sub-advisors does not affect the performance of incumbent sub-advisors. Overall, there is no evidence that sub-advisor selection decisions by fund families benefit sub-advised MF's performance. Outperforming sub-advisors with larger style drift are less likely to be hired, and the more a sub-advisor deviates from its investment mandate, the more likely it is to be fired.

Keywords: Mutual Funds, sub-advisors, managerial turnover, return-chasing, competition.

JEL classification:

* All errors are ours.

3.1 Introduction

Mutual fund families often delegate or outsource the asset management of a mutual fund to sub-advisors rather than managing the assets in-house. Prior empirical analysis of mutual funds has established an absence of persistence in generating abnormal returns and a convex flow-performance relationship, suggesting that investors chase the best-performing funds.[1] Given that mutual funds that outsource management are typically more sophisticated, knowledgeable, and better resourced than the average mutual fund investor, one might expect fund families to do a better job in picking outperforming managers. The large number of sub-advisor hiring and firing decisions made by fund families allows us to test this hypothesis. Between January 1995 and December 2016, we count 1,239 hiring and 809 firing decisions made by fund families, across 426 sub-advised mutual funds. Fund families may choose to replace a sub-advisor for "good" reasons, such as to improve fund returns and to lower risk, or for "bad" ones, such as to increase fund flows and asset management revenues despite any improvement in performance.

This paper makes three main contributions. First, we analyse what drives the large number of hiring and firing events and their resulting impact on mutual fund and sub-advisor returns. Second, we study competition among sub-advisors by exploiting the fact that the existence of sub-advised (or outsourced) funds with two or more sub-advisors effectively introduces competition into the management of these funds. The presence of more than one sub-advisor leads to a split of management fees among sub-advisors, which in turn encourages greater monitoring between them and appears to benefit performance (Moreno et al. (2018)). Exploiting the fact that 32% of the outsourced funds in our sample are managed by two or more sub-advisors, we examine if the performance of the incumbent sub-advisors varies when either one of them is fired or a new sub-advisor is hired and if this performance is different from that of the recently hired/fired sub-advisors. Finally, we study how actively the fund family monitors its sub-advisors beyond looking at performance and whether deviating from the investment mandate affects a subadvisor's chances of being hired or fired. Despite not being able to outsource asset

[1] See, for example, Grinblatt and Titman (1992), Brown and Goetzmann (1995),Elton et al. (1996), Carhart (1997), Daniel et al. (1997), Sirri and Tufano (1998), and Berk and Green (2004).

management to improve returns, advisors may be able to act as monitors simply to ensure that sub-advisors follow the mandates they are given.

Our results show that advisors exhibit similar behaviour to retail investors when selecting a mutual fund (Sirri and Tufano (1998)). Fund families tend to fire sub-advisors following underperformance and hire them after outperformance. When the fund itself is outperforming, there are fewer hirings and firings. Furthermore, such changes do not lead to higher returns, lower volatility, better Sharpe ratios, and higher asset flows. We also do not find that competition among sub-advisors improve performance. Turnover does not affect the risk-adjusted returns of incumbent sub-advisors, and their subsequent performance is not different from those of recently hired/fired sub-advisors. Finally, the likelihood that a sub-advisor is fired increases with its degree of style drift (measured by differences of factor loadings relative to its Morningstar style's average) and that the likelihood of being hired following outperformance is reduced by the style drift.

We focus on all sub-advised U.S. domestic equity funds available on Morningstar between 1995 and 2016. We collect EDGAR data on funds and their sub-advisors, such as the date sub-advisors' are hired or fired, the identity of fund families, among others. On average, 7.8% of domestic equity funds are sub-advised, but the proportion has been increasing over time and in 2016 it reached 11.2% of the total. Unlike previous work, we collect sub-advisors' return data from Morningstar and the Informa PSN Separately Managed Accounts database, which allows us to extend the literature by studying performance at the sub-advisor level instead of only at the fund level. We match the sub-advisor name and investment style reported on EDGAR with the mutual fund in Morningstar which has the same portfolio manager and investment style to estimate the pre-hiring performance of a sub-advisor. In total, we study 426 funds that have more than 382 unique sub-advisors.

Funds experiencing persistently poor performance face the threat of fund outflows and lower management fees.[2] If a fund family has the ability to identify future performance based on past performance, we would expect that the decision to hire or fire a sub-advisor is related to

[2] Past research shows that retail and institutional investors chase performance when selecting fund managers (e.g., Sirri and Tufano (1998), Karceski (2002), Knittel et al. (2004), and Goyal and Wahal (2008)).

past performance. On the one hand, such decisions send a positive signal to the market about the fund family's selection ability. In such a case, more capital is directed towards the fund, increasing the asset management company's fees. On the other hand, if manager performance is not persistent, underperforming funds may decide to change sub-advisors because fund investors make decisions based on past performance. Therefore, regardless of the ability of fund families to select sub-advisors and of the existence of performance persistence, we would expect that sub-advisor turnover depends on the past performance of the fund and of the sub-advisor. Similar to previous literature, our results show that fund families chase past performance in selecting/terminating sub-advisors (Goyal and Wahal (2008) and Kostovetsky and Warner (2015)). This behaviour is similar to the one found for retail and institutional investors.

A natural question is if these sub-advisor changes can ultimately improve the performance or risk profile of the fund. There are at least three different channels through which performance could be enhanced. The first is that, by hiring past winners, the out-performance of the chosen sub-advisor continues, improving the fund's future returns. However, we find that the performance of the fund following sub-advisor changes is in fact worse, mostly due to the mean-reversion of sub-advisor returns after being hired/fired. Investors could still be better off if the risk of the fund is reduced, but there is no significant reduction in fund volatility or Sharpe ratio after sub-advisor changes. Finally, in spite of any improvements to clients, the fund family might still benefit by changing sub-advisors if it leads to higher asset flows. Yet again, our results do not show any significant effect due to sub-advisor changes on asset flows.

The effect of competition among sub-advisors on fund performance can be studied through the lens of multi-advised funds. Fund performance depends not only on the returns of the sub-advisor that is hired or fired. but also on any effects on incumbent sub-advisors that already manage a portion of the fund's assets. Therefore, a second channel that might improve the performance of funds that already outsource asset management is through competition. The arrival of a newly hired sub-advisor may increase competition and lead to better performance of the incumbent sub-advisors. In turn, this could improve overall fund performance, even as the abnormal performance of the newly-hired sub-advisor decreases after being hired (e.g.

Goyal and Wahal (2008)), if the net effect is positive. Similarly, the termination of an under-performing manager might discipline the remaining sub-advisors, improving fund performance. Using unique data on the returns of individual sub-advisors of a fund, we find that performance of the incumbent sub-advisors does not improve in the 36-month period after the change. A third channel is that the performance of a newly-hired sub-advisor may still outpace that of the incumbents, with a net positive impact on abnormal returns. Similarly, the performance of a fired sub-advisor may be relatively worse than that of the remaining incumbents, resulting in higher performance after the sub-advisor's departure. Our results do not find any difference in performance between newly-hired and existing sub-advisors. In short, our empirical analysis provides no evidence that changing the level of competition through variation in the number of sub-advisors of a fund improve its performance.

Finally, we study whether deviating from the investment mandate affects a sub-advisor's chances of being hired or fired. Sub-advisors possess some degree of freedom to drift away from their mandates and deviations from a mandate might be used to exploit performance-enhancing stock selection and market-timing opportunities, increasing fund performance (Chevalier and Ellison (1999)). Alternatively, managers may deviate from their mandates to increase fund flows and, ultimately, their own compensation and that of the fund family. **?** show that funds with higher levels of style volatility underperform, suggesting that style drift might be motivated by agency issues. Relatedly, **?** finds that funds that change the volatility of their portfolios more often (i.e., higher "risk shifting") also underperform. These results lead us to hypothesize that sub-advisors with larger style-drift from their mandates are less likely to be hired and more likely to be fired. Our estimates show that fund families are less likely to hire outperforming sub-advisors if they exhibit larger style drift. When firing a manager, excessive style drift of sub-advisors is associated with a higher likelihood of being fired, regardless of its performance. These results show that fund advisors actively monitor sub-advisors and use high-powered incentives to make sure that they stay in line with the stated investment style, consistent with Chen et al. (2013) and Chevalier and Ellison (1999)

3.2 Hypotheses Development

Asset managers aim to maximize profits extracted from fees based on total assets under management. For any sub-advisor turnover decision to be rational, the asset manager should benefit from a growing asset base post turnover through either improved performance or incremental flows. Below, we describe four testable hypotheses that investigate the different ways in which asset managers' choice of sub-advisors is related to fund and sub-advisor characteristics.

Mutual fund investors tend to chase past performance and allocate capital to outperforming funds (Sirri and Tufano (1998) and Karceski (2002)). As a consequence, we expect asset managers aiming to expand their asset base to turnover sub-advisors if the fund suffers from persistent underperformance or consistent outflows. On the other hand, management turnover can be costly, and searching for new sub-advisors requires considerable skill and effort. Therefore, if a fund is outperforming, there should be little incentive to replace sub-advisors and the likelihood of turnover would be negatively related to the fund's performance. Previous literature documents similar patterns for both internal fund manager turnover and external sub-advisor departure. Chevalier and Ellison (1999) find a negative relation between fund past performance and the likelihood of management turnover. Kostovetsky and Warner (2015) find an inverse relation between sub-advisor departure and past fund flows. Such return-chasing behaviour is documented not only among retail investors, but also among institutional investors (Knittel et al. (2004) and Goyal and Wahal (2008)). Therefore, advisors are likely to engage in return-chasing behaviour and their decision to hire and fire a sub-advisor should be affected by past performance, regardless of whether mutual fund performance is persistent or not. This leads to our first hypothesis:

Hypothesis 1 *The turnover of a sub-advisor is related to the past performance of the sub-advisor and of the fund itself: hiring (firing) a sub-advisor is more (less) likely if the sub-advisor is outperforming and less (more) likely if the fund is outperforming.*

Mutual fund families have access to more expertise and resources than retail investors to research the investment skill of other fund managers (Guercio and Tkac (2002)). As such, they are expected to make smarter decisions in delegating asset management. If asset managers are skilled at identifying outstanding sub-advisors, we expect an improvement in the fund's performance, either via higher returns or lower risk, following sub-advisor turnover. Furthermore, given the convex relationship between performance and fund flows (Sirri and Tufano (1998)), we expect fund families to incorporate retail investor's preference for mutual funds with higher past performance, replacing sub-advisors in expectation of higher flows. This hypothesis is summarized below:

Hypothesis 2 *Following sub-advisor turnover, the risk and return characteristics of the fund improves. The fund should also experience higher inflows of capital.*

Our data allow us to decompose what drives the change in performance of multi-advised fund following sub-advisor turnover. Consider the case where an outperforming sub-advisor is hired by an existing outsourced mutual fund managed by an existing sub-advisor. There are three channels through which fund performance could be affected by sub-advisor turnover in this setting. First, the pre-hiring outperformance of the newly-hired sub-advisor persists after hiring and improves the fund's return. Second, the hiring of a new sub-advisor improves the performance of the incumbent sub-advisors via increased sub-advisor competition. This may be the case even if the new sub-advisor's outperformance subsequently reverts to the mean (Goyal and Wahal (2008)). Notwithstanding an expectation of mean reversion in post-hiring sub-advisor performance, it is still rational for a fund to hire a new sub-advisor if the increased competition among sub-advisors forces the incumbent ones to perform better, as long as the net effect on fund performance is positive. Third, notwithstanding mean-reversion in post-hiring sub-advisor performance, the post-hiring performance of the newly-hired sub-advisor might still exceed that of the incumbents. Our third hypothesis focuses on the impacts of managerial turnover on incumbent sub-advisors:

Hypothesis 3 *When a fund family adds one or more sub-advisors to a sub-advised fund, the incumbent sub-advisor(s) exhibit outperformance post-hiring as competition between sub-advisors increases. Similarly, incumbent sub-advisors underperform after firing as competition recedes.*

Advisors state the objective and style ("the mandate") of the mutual fund that guides investors on how they intend to invest their capital. Similarly, sub-advisors are in turn also guided in their investment decisions by the same fund mandate chosen by the fund advisor. Sub-advisors possess some degree of freedom to drift away from the fund mandate and deviations from a mandate might be used to exploit performance-enhancing stock selection and market-timing opportunities, increasing fund performance (Chevalier and Ellison (1999)). Alternatively, managers may deviate from their mandates to increase fund flows and, ultimately, their own compensation. **?** show that funds with higher levels of style volatility underperform, suggesting that style drift might be motivated by agency issues. Relatedly, **?** find that funds that change the volatility of their portfolios more often (i.e., higher "risk shifting") also underperform. Since we believe that fund advisors are better monitors than ordinary mutual fund investors, we expect that sub-advisors with larger deviations from their mandates are less likely to be hired and more likely to be fired. Our last hypothesis is summarized below:

Hypothesis 4 *Deviations from its investment mandate decrease the likelihood that a sub-advisor is hired and increase the likelihood that a sub-advisor is fired.*

3.3 Data and Descriptive Statistics

3.3.1 Sample Selection

First, we download all sub-advised U.S. domestic equity funds from Morningstar. After removing those with duplicate share classes, the initial sample comprises 1,061 unique funds from January 1995 to December 2016. Our sample excludes owner funds that do not have independent subadvisors or advisors that are affiliated with the main advisor (306 funds), index funds (240), those without return and fee data (36), and funds that were reorganized or merged with other companies (8 funds). For the final sample of funds, we download information on assets under management, inception dates, fund manager names, investment style, returns, and expense ratios.

From SEC fillings on EDGAR, we collect data on the identity of the sub-advisors and the date a sub-advisor is hired or fired. Our data also include information on sub-advisor firms, such as inception dates and their total assets under management at the time of hiring. SEC fillings also disclose when sub-advisors are direct subsidiaries of the main advisor or trust of the fund.

For each fund sub-advisor, we obtain return data around hiring and firing events from Morningstar and the Informa PSN Separately Managed Accounts database. In order to obtain pre-hiring performance of a sub-advisor, we match the sub-advisor name and investment style reported on EDGAR with the mutual fund in Morningstar which has the same portfolio manager and investment style. We then select the class with the longest return history available, using the institutional share class whenever possible. Where we cannot find a match in Morningstar, we search on the Informa PSN database. In the case of 25 sub-advisors without an exact match using this procedure, we select the returns for the closest available strategy on Morningstar (e.g., Large Cap Growth) as long as the fund managers' names are the same.

3.3.2 Variable Definitions

Return Performance

We study the effects of hiring and firing of sub-advisors by fund advisors in event time and define the event date as the month in which the hiring or firing takes place. We estimate returns at both the fund level and the sub-advisor level for 12, 24, and 36 month windows around each event. We evaluate performance up to three years to account for long-term mean reversion following (Goyal and Wahal, 2008). In the remaining analysis, we denote the event window by close intervals in months.

Abnormal returns are computed in two alternative ways. The first is based on Carhart's (1997) four-factor abnormal returns computed using a 36-month rolling window:

$$\alpha_{i,t} = R_{i,t} - R_{F,t} - \hat{b}_{i,t-1} RMRF_t - \hat{s}_{i,t-1} SMB_t - \hat{h}_{i,t-1} HML_t - \hat{p}_{i,t-1} PR1YR_t, \tag{3.1}$$

where $R_{i,t}$ is the excess return of portfolio i in month t over the one-month T-bill rate $(R_{F,t})$

and $RMRF_t$ is the excess return on a value-weighted aggregate market proxy in month t. SMB_t, HML_t, and $PR1YR_t$ are returns on value-weighted, zero-investment, factor-mimicking portfolios for, respectively, size, book-to-market equity, and one-year return momentum.[3]

The second abnormal return measure is based on a portfolio's excess returns relative to its benchmark (BMK). At the fund level, we use the fund's net monthly returns relative to its benchmark. Where there is only a single sub-advisor, the returns of the fund and the sub-advisor are identical. When there is more than one sub-advisor, we obtain the performance of each sub-advisor by using the net monthly return of the matched portfolio for each sub-advisor. Information on the primary prospectus benchmark and the monthly returns on the benchmark indices are collected from Morningstar. If the benchmark information is missing, we replace it with the most popular benchmark used by funds within the same 3x3 Morningstar style category. A list of the categories along with the corresponding default benchmarks is provided in Table A3.1.[4] Alphas are estimated using the following equation:

$$R_{i,t} - R_{F,t} = \alpha_i + \beta_{i,t}(R_{BMK_i,t} - R_{F,t}) + \varepsilon_{i,t}, \tag{3.2}$$

where $R_{i,t}$ denotes the net monthly return for portfolio i during month t and $(R_{BMK_i,t} - R_{F,t})$ denotes the excess return of the primary benchmark over the risk-free rate for fund i in month t.

Other Performance Characteristics

We compute the Sharpe Ratio (SR) and Information Ratio (IR) for up to three years before and after the hiring and firing decisions. We calculate the Sharpe Ratio at time t of an investment manager i over H periods as:

$$SR_i(t, H) = \frac{\overline{CAR_i(t, H)}}{\sigma_{AR}}, \tag{3.3}$$

[3]Monthly factor returns on market, size, value, and momentum factors come from Ken French's website, accessed through https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[4]Soggiu et al. (2020) find that a mismatch may exist between the benchmark disclosed in the prospectus and the fund's mandate. In such cases, assigning a category benchmark that is better aligned to the fund's objective produces more reliable measures of benchmark-adjusted performance.

where the cumulative abnormal return (CAR) is defined as the arithmetic sum of the monthly abnormal returns (ARs) over the risk-free rate, i.e.:

$$CAR_i(t,H) = \sum_{s=t}^{t+H-1} (R_{i,s} - R_{F,s}), \tag{3.4}$$

and σ_{AR} denotes the standard deviation of the abnormal returns over H periods. The Information Ratio is calculated similarly, with cumulative abnormal returns replaced by the cumulative benchmark-adjusted excess returns.[5]

Following Chevalier and Ellison (1999), we compute two measures of risk-taking deviations. The first is a fund's idiosyncratic risk, defined as the standard deviation of residuals from the factor model regression on monthly returns. The second measure examines style drift (*Beta Dev*), defined as the square root of the sum of squared deviations of a fund's size and value betas relative to the respective betas of its 3x3 Morningstar category:[6]

$$Beta\ Deviation_{i,t} = \sqrt{(\beta_{size_{i,t}} - \overline{\beta_{size,t}})^2 + (\beta_{value_{i,t}} - \overline{\beta_{value,t}})^2}. \tag{3.5}$$

Other Variables

In addition to the risk and return metrics, we also examine two measures of fund flow. The dollar flow in a given month is defined as:

$$Dollar\ Flow_{i,t} = TNA_{i,t} - TNA_{i,t-1} \times (1 + r_{i,t}), \tag{3.6}$$

where $TNA_{i,t}$ denotes the total net assets of fund i at the end of month t, and $r_{i,t}$ denotes the gross return of fund i during month t. Percentage flow is equal to Dollar Flow scaled by the

[5]The Information Ratio measures the skill of a portfolio manager at generating returns in excess of a given benchmark, while incorporating the consistency of the performance through the tracking error in the denominator.

[6]The investment mandate of each fund falls into one of the 3x3 categories defined on the size and value dimensions, as shown in Table A3.1. Therefore, style drift of a fund could be more accurately described by the beta deviations on the size and value factors net of any effect from the market and momentum factors.

start-of-period TNA:

$$\%Flow_{i,t} = \frac{Dollar\ Flow_{i,t}}{TNA_{i,t-1}}. \tag{3.7}$$

Fund flow is highly sensitive to past performance (Sirri and Tufano, 1998) and to other characteristics such as fund age and total fund size (Kostovetsky and Warner, 2015). Following Kostovetsky and Warner (2015), we estimate abnormal flow alpha as follows:

$$Flow\ Alpha_{i,t} = Dollar\ Flow_{i,t} - \left(\hat{b}_{i,t-1}\overline{R_{i,pr1yr}} + \hat{a}_{i,t-1}Age_{i,t} + \hat{s}_{i,t-1}TNA_{i,t}\right), \tag{3.8}$$

where $\overline{R_{i,pr1yr}}$ denotes the prior-12-month average of fund i's net return, fund age and fund TNA and the coefficients $\hat{b}_{i,t-1}, \hat{a}_{i,t-1}, \hat{s}_{i,t-1}$ are estimated using data from the previous 12 months.

3.3.3 Descriptive Statistics

Table 3.1 summarizes the descriptive statistics of our final sample. In Panel A, we report characteristics of the 426 mutual funds outsourced to 382 sub-advisors and distributed by 84 different fund families. A sub-advised fund on average has \$491.8m AUM with monthly four-factor alpha of -0.111% and benchmark alpha of -0.059%. The average monthly percentage flow is 0.3%. The median and mean number of sub-advisors per fund is 1 and 2.62 respectively, while the median and mean number of mutual funds managed by the same sub-advisor at the same time is 1 and 3.25 respectively. The median tenure for a sub-advisor is 57 months, and the mean is 71. In Panel B, we identify 1,239 cases of a sub-advisor being hired and 809 cases of being fired. Among these hirings, there are 50 instances of hiring upon fund inception, whilst among the firings there are 18 instances where sub-advisory contracts end with the delisting of the fund. Less than 10% of the hiring events (105) arise when fund families move from in-house management to outsourcing, while more than a quarter of the firing events (229) result from outsourced funds shifting back to in-house management. Finally, in Panel C, we summarize the 1,291 fund-month observations where there are sub-advisor turnovers, comprising 591 pure hirings, 374 pure firings and 326 cases of simultaneous hiring and firing.

Table 3.2 presents the percentage of sub-advised funds covered in the sample, together with the distribution over the number of sub-advisors for each year from 1995 to 2016. The number of sub-advised funds grow over time, reaching 350 at peak time in 2015, which represents 11.24% of all US domestic equity MFs. On average, 68.16% sub-advised funds have only one single sub-advisor, 10.45% sub-advised funds have two, and another 6.47% sub-advised funds have three. The other 14.92% sub-advised funds have four or more sub-advisors on average.

3.4 Results

3.4.1 Impact of Past Performance on Sub-advisor Turnover

In this section, we investigate the determinants of sub-advisor turnover decisions by mutual fund families to test the hypotheses laid out in Section 2. We start by examining the relationship between sub-advisor hiring/firing decisions and past performance of both the fund and the sub-advisor. We do so by estimating the following panel probit regression:

$$\begin{aligned}\Pr(Hired_{i,t}/Fired_{i,t}=1)= \quad & \Phi(\mu+\beta_1 \textit{Sub-advisor Performance}_{i,t-1} \\ & +\beta_2 \textit{Fund Performance}_{i,t-1}+\lambda' X_{t-1}+\gamma_t+\varepsilon_{i,t}),\end{aligned} \tag{3.9}$$

where $Hired_{i,t}/Fired_{i,t}$ are indicator variables that take a value of one if sub-advisor i is hired/fired by the fund in month t and zero otherwise. The notation $\Phi(\cdot)$ stands for the cumulative distribution function of a standard normal distribution. Our main explanatory variables are performance measures computed for 12, 24, and 36 month-windows pre-hiring/firing, for both the fund and the sub-advisor, based on the Carhart's (1997) four-factor abnormal return (*Factor Alpha*) and the benchmark-adjusted abnormal return (*Benchmark Alpha*). Control variables are lagged one month and represented by the matrix X_{t-1}, including *Log(Fund TNA*, *Log(Sub-advisor Age)*, *Log(Fund Age)* and *No. of Funds in Family*. For firing decisions, we include *Log(Sub-advisor Tenure)* and an indicator variable *Co-branding*, equal to 1 if the sub-advisor's name appears in the fund's name, zero otherwise as in Moreno et al. (2018). γ_t captures time fixed effects. Standard errors are reported in brackets and clustered at the fund-level.

We report results in Table 3.3. At the fund level, the past performance is associated with changes in the likelihood of sub-advisor turnover. Coefficients on the prior 12-month *Factor Alpha* and *Benchmark Alpha* performance measures are both negative and statistically significant at the 1% level in columns (1)-(4), for both hiring and firing. These results imply that that underperformance at the fund-level drives sub-advisor turnover and consistent with Chevalier and Ellison (1999), who show that closures of outsourced funds are more sensitive to low past performance than in-house funds. At the sub-advisor level, outperforming sub-advisors are more likely to be hired. Coefficients on both sub-advisor factor alpha and benchmark alpha are positive and significant across columns (1)-(2). The statistical significance can be dated as far back as three years pre-hiring. Similarly, results from column (3)-(4) show that sub-advisors are more likely to be terminated for persistent under-performance. Coefficients on both the sub-advisor factor alpha and benchmark alpha measures are negative and statistically significant for at least two years before firing. They show that outperforming sub-advisors are more likely to be hired by a fund and less likely to be fired.

TABLE 3.3 ABOUT HERE

When observing the results for control variables, most are not statistically significant, but there are some exceptions. First, younger funds appear to be more aggressive in firing underperforming sub-advisors. Second, fund families with fewer fund offerings are more likely to hire new sub-advisors. This is partly due to the lack of relevant expertise within the family to manage the portfolio in-house (Debaere and Evans, 2014). Third, sub-advisors with relatively longer tenure are more likely to be fired.

Taken together, these findings support Hypothesis 1 and suggest that fund families use past performance in their decision to hire/fire sub-advisors. Both the absolute performance measured by Carhart's alpha and the relative performance measured by benchmark alphas at the sub-advisor level matter, and the effect can persist up to three years. The return-chasing behaviour of sub-advised mutual funds is similar to that exhibited by both retail investors and plan sponsors in picking investment managers (Sirri and Tufano (1998), Bergstresser and Poterba (2002), and Sapp and Tiwari (2004), and Goyal and Wahal (2008)).

3.4.2 Impact of Sub-advisor Turnover on the Returns, Risk, and Flows of a Fund

In this section, we examine if fund advisors have superior skills when hiring/firing sub-advisors. We construct a balanced sample by excluding fund observations without continuous return data throughout the event window. In this way, the cases where sub-advisor hiring/firing changes coincide with a fund's launch or delisting are excluded from the sample. We also remove events in which hiring a sub-advisor would make the fund go from being managed in-house to being outsourced (and vice versa for firing). This ensures that our results are not due to the effects from outsourcing and allows us to focus on the increase/decrease in competition among sub-advisors.[7]

For an initial picture on how fund performance changes in the months around sub-advisor turnover events, Figure 3.1 plots the annualized sample means of factor alphas for sub-advisors (top panel), for all turnovers (middle panel), and for three different turnover types separately (bottom panel). On the top panel, we can see that hired sub-advisors are outperforming in the previous 36 months, while those fired have been underperforming. From the middle panel, we can infer that there is no difference in performance before or after a change in sub-advisors. The bottom panel decomposes these events and shows there is no difference in performance regardless of the type of event (pure hiring, pure firing, or mixed).

FIGURE 3.1 ABOUT HERE

For a more formal analysis, we employ an event study framework to compare fund perfor-

[7] Chen et al. (2013) show that outsourced funds underperform the in-house managed funds by about 52 basis points per year. Chuprinin et al. (2015) find that in companies running both outsourced and in-house funds, in-house funds outperform outsourced funds by 0.85% annually, which amounts to 57% of the expense ratio.

mance before and after sub-advisor turnover:

$$\begin{aligned} Fund\ Characteristic_{i,t} = \quad & \mu + \gamma_1 Post\ Turnover_t * Pure\ Hiring_i \\ & + \gamma_2 Post\ Turnover_t * Pure\ Firing_i + \lambda Post\ Turnover_t + \varepsilon_{i,t}. \end{aligned} \tag{3.10}$$

Fund Characteristic denote the different dependent variables, including Carhart's (1997) four-factor alpha, benchmark alphas, Sharpe ratios, volatility, and asset flows for *t=12*, *24*, and *36*-month windows around hiring/firing events. *Post Turnover* is an indicator variable that equals one for post-event months and zero otherwise. *Pure Hiring* and *Pure Firing* are indicator variables taking a value of one if a fund only, respectively, hires or fires a sub-advisor within a given month and zero otherwise (i.e., when a fund does not simultaneously hire a sub-advisor and fire another). The interaction terms of these variables with *Post Turnover* allow us to perform a difference-in-difference analyses relative to the mixed hiring&firing before turnover as the baseline case without any control variables. The parameter γ_1 and γ_2 capture the marginal effect of, respectively, a pure hiring and a pure firing event on future fund performance relative to the performance of a fund with a mixed hiring&firing before the event date.

It would make sense for advisors to outsource asset management based on past performance, incurring the associated costs of the change, if the fund's future performance or asset flows improve. On Panel A in Table 3.4, the estimated coefficients for the constant in columns (1)-(3) show that a fund's Carhart's (1997) before changing sub-advisors in mixed hiring&firing events is negative and statistically significant regardless of the event window, varying from -0.159% to -0.140% per month. After mixed hiring&firing events, there are no improvements to fund alpha for the 12-month and 24-month windows (i.e., the *Post Turnover* coefficients are not statistically significant). *Pure Hiring* and *Pure Firing* events are not associated with any marginal improvements relative to the baseline mixed hiring&firing changes after changes in sub-advisors. Results using the benchmark alpha in columns (4)-(6) are similar. Columns (7)-(9) show a significant improvement in the Sharpe ratio two and three years post turnover. Relative to a baseline mixed hiring&firing event, the average Sharpe ratio of a fund increases from 0.114 to 0.114 + 0.067 = 0.181 in the two-year event window in column (8) and more than triples in the three-year event window (from 0.055 pre-turnover to $0.055 + 0.123 = 0.178$ post-turnover).

TABLE 3.4 ABOUT HERE

The increase in Sharpe ratios of a fund can be explained by a significant decrease in fund volatility after sub-advisor changes. In columns (1)-(3) of Panel B, we observe a negative and statistically significant decrease in volatility post turnover. In columns (2)-(3), we find a reduction of almost 10% in a fund's return volatility relative to the baseline. There is no statistical difference after pure hiring and pure firing events.

An alternative to improving fund performance that can justify sub-advisor changes is that it increases asset flows, leading to higher management fees to the fund family. We examine the effect of sub-advisor changes on flow (columns (4)-(6)) and abnormal flow (columns (7)-(9)). We do these findings indicate that although sub-advisor turnovers do not improve future fund returns and asset inflow directly, they may lower the fund's aggregate risk-taking in the long-run, and hence provide investors with a better risk-return profile. However, these results may be affected by mean reversion and other fund-specific and time-specific variation.

Controlling for Past Fund Characteristics

In Table 3.5 we estimate the following fixed-effect panel regression:

$$\begin{aligned} Fund\ Characteristic_{i,t} = \quad & \mu + \beta_1 PureHiring + \beta_2 PureFiring + \beta_3 Hiring\&Firing \\ & + \lambda Fund\ Performance_{i,t-1} + \delta' Controls_{i,t-1} + \gamma_i + \gamma_t + \varepsilon_{i,t}, \end{aligned} \tag{3.11}$$

where the dependent variable *Fund Characteristic* denotes Carhart's (1997) alpha, benchmark alphas, and Sharpe ratios measured for the fund i at time t over the next 12, 24, and 36-month periods. Performance measures for the previous values for each of these windows are included to capture mean reversion in performance. We also include fund and time fixed effects (γ_i and γ_t). Standard errors are clustered at the fund-level to account for unobserved heterogeneity across different fund advisors. The same set of control variables as in Table 3.3 is included, the time-invariant fund-specific characteristics are absorbed into fund fixed effects.

Table 3.5 About Here

We find a negative and significant relation between *Log*(*Fund TNA*) and fund future performance reflects the decreasing return to scale among asset managers (Pástor et al. (2015)). Fund age is also negatively related to fund performance, suggesting younger funds who have better incentives tend to do better. Results also show substantial mean reversion in sub-advised funds' performance after managerial turnover. In columns (1)-(3), we can see that Carhart's (1997) alpha over the next three years are negatively correlated to past alphas (*Fund Factor Alpha*), and the relationship remains significant at the 1% level for all three years. In columns (4)-(6), we find that future benchmark alphas are also negatively related to past 12-month benchmark alphas. The mean-reversion for the Sharpe ratios, shown in columns (7)-(9), is weaker, with only the fund's Sharpe ratio in the past 12-month being negative and statistically significant.

More importantly, coefficients on the three sub-advisor turnover indicators — *Pure Hiring*, *Pure Firing*, and *Hiring&Firing* — captures how the future performance of a fund is affected by sub-advisor turnover. This represents the net effect related to sub-advisor turnover, controlling for mean reversion. Across all performance measures and time windows, coefficients on sub-advisor turnover indicators consistently point to the fact that changing sub-advisors does not improve a fund's future performance metrics. In the case of *Pure Firing* and *Hiring&Firing* the effects are in fact negative, hurting the fund. The Sharpe Ratios improvements found in Panel A of Table 3.4 disappear after we include controls and fixed effects, rejecting the first part of Hypothesis 2. In columns (1)-(3) of Table 3.6 we can see that sub-advisor changes are also unrelated to fund volatility, reinforcing the findings for the Sharpe ratio.

An alternative reason for fund families to change sub-advisors could be that if it increases flows into the mutual fund, even if hurts fund performance. For the decision to replace sub-advisors to be rational. the expected benefit to the fund advisor must exceed the associated costs. In this section, we investigate if sub-advisor turnover could assist future flows.

Table 3.6 reports the results of how sub-advisor changes (i.e., pure hirings, pure firings, and mixed hirings/firings) affect a fund's percentage monthly flows (columns (4)-(6)) and flow alpha (columns (7)-(9)) in the 12, 24, and 36 months after turnover. The positive and significant coefficients on the prior 12 and 24-month percentage flows in Panel A show substantial persistence of flows over time. After controlling for this effect, general sub-advisor turnovers appear to be less harmful to future percentage flows. Pure hiring events improve future percentage flows in the short-run, with a 0.41% increase in the first year and 0.22% in the following two years, but the effect is not significant in the 36-month window. However, unlike Kostovetsky and Warner (2015) there is no significant improvement in flows after pure firing and mixed hiring/firing events.[8] In columns (7)-(9), we repeat the analysis using the future flow alpha, finding that the Pure Hiring coefficient is no longer statistically significant. Overall, changing sub-advisors does not increase flows, rejecting the final part of Hypothesis 2.

TABLE 3.6 ABOUT HERE

Finally, the effect of sub-advisor changes (*Turnover*) might show considerable cross-sectional variations. On Table A3.3 of the Appendix, we test if the impact of sub-advisor turnover on *future* performance and flow measures vary conditional on two specific fund-level characteristics: the size of a fund and the fraction of its total net assets relative to the fund family's total net assets. First, columns (1)-(3) of Panel A show that size brings economies of scale and expertise in picking sub-advisors. Consistent with this argument, we find that sub-advisors' changes made by larger fund advisors are significantly less harmful to post-turnover fund abnormal performance than those made by smaller funds. Second, in columns (4)-(6) we test if the higher share of a fund on its family's total assets may increase performance, due to more attention and resources being directed to the fund. The interaction term coefficients are positive, but not statistically significant, using a 24 and 36 event-windows. In Panel B of Table A3.3, we test if the past performance of a fund's alpha affect changes the impact

[8]In unreported tests, we run the fixed effect panel regression on the subsample using the same period as Kostovetsky and Warner (2015), the empirical pattern is consistent with the whole sample.

of sub-advisor changes on *future* asset flows. Columns (1)-(3) show that a fund's alpha does not change how sub-advisor changes affect the flow alpha. In columns (4)-(6), the *Turnover* coefficient is positive and significant at the 10% level for the 12 and the 24-month window. The *Turnover*Fund Flow Alpha[-12,-1]* is negative and significant at the 10%level. These results imply that changing sub-advisors may reduce *future* abnormal flows if the fund's *past* abnormal alpha are large enough.

3.4.3 Impact of Sub-advisor Turnover on Sub-advisors' Performance

In the previous section, we examine fund characteristics before and after sub-advisor turnover. Results suggest that advisors have no timing ability to select sub-advisors, and that these changes in fact hurt future performance and do not improve asset flows. For instance, in the top panel of Figure 3.1, we plot the annualized sub-advisor Carhart's (1997) alpha for 12, 24, and 36 months before and after hiring and firing. We can clearly observe that sub-advisor abnormal performance is mean-reverting after being hired and fired. However, the middle and bottom panel shows that the overall effect on the fund's return is insignificant.

FIGURE 3.1 ABOUT HERE

Since the return of a sub-advised fund is the value-weighted average of all of its sub-advisors' individual returns, in this section we study how the performance of the individual sub-advisors within a fund are affected after sub-advisor turnover. Hypothesis 3 discusses how competition among sub-advisors can affect sub-advisors' performance and the overall returns of the fund through three channels. For example, suppose that a sub-advised fund hires an outperforming sub-advisor. The first channel is that by picking a past winner, its outperformance persists, positively affecting the returns of the fund. The second possibility is that the newly-hired sub-advisor increases the performance of *existing* sub-advisors through increased competition. Therefore, overall fund performance may improve by changing incentives for incumbent sub-advisors. The third channel is that, even accounting for the underperformance after hiring a sub-advisor (Goyal and Wahal (2008)), the overall performance of the new sub-advisor is still relatively better than that of incumbent ones.

In Panel A of Table 3.7 we examine the abnormal return persistence of *sub-advisors*, rather than the fund's, before and after sub-advisor turnover. We only include sub-advisors with continuous return data around the symmetric event windows to create a balanced sample, excluding any who are fired within 12, 24, and 36-month windows post-hiring. In columns (1)-(3), we find that the sample mean for sub-advisor's Carhart's (1997) alpha prior to hiring varies between 0.085% using the 12-month window and 0.166% using the 36-month window. This shows that sub-advisors are outperforming immediately before being hired. Conversely, for firing events in columns (4)-(6), we find that sub-advisors o average underperform before being fired. For example, in column (4) we find that the average sub-advisor alpha in the 12 months prior to being fired is equal to -0.177%. However, our results also show substantial mean reversion of sub-advisors' returns across all event windows. For example, in column (1) we find that the sub-advisor's factor alpha drop significantly from 0.085% per month to $-0.011\%(= 0.085\% - 0.096\%)$ in the 12 months after being hired. The decrease in performance remains indistinguishable from zero in the 36-month window. Similarly, after being fired, sub-advisors' factor alphas improve gradually from -0.177% to $-0.11\% = (-0.177\% + 0.067\%)$ in the first year and -0.051% after three years.

Finally, sub-advisors hired by funds that already have multiple sub-advisors may face steeper competition. Therefore, we add the variable *Single/Multi*, which denotes a fund that changes from having a single sub-advisor to having multiple ones (or vice-versa) and interact it with the *Post Hiring/Firing* indicator. In columns (1)-(3), we do not find consistent evidence that funds that go from having no competition between sub-advisors to being multi-advised in the long-run have any difference in performance, although in the first year we find that performance is in fact worse. We find a similar lack of impact in columns (4)-(6), where we investigate funds that go from having multiple sub-advisors to a single one, reducing the competition. In columns (7)-(12) we use the benchmark alpha as the dependent variable, finding similar conclusions.

Table 3.7 About Here

In Panel B of Table 3.7, we examine the disciplinary effect of sub-advisor turnover on the incumbent sub-advisors. We expect that the introduction of a new sub-advisor would boost incumbent performance through increased competition. Similarly, firing a poorly performing sub-advisor may discipline the remaining sub-advisors in the fund. We identify incumbent sub-advisors of a fund 12, 24, and 36 months before the sub-advisor change that also remain as sub-advisors for similar periods thereafter. Dependent variables are the value-weighted alphas across all incumbents around the sub-advisor turnover. Coefficients on the *Post Turnover* indicator are consistently negative across both performance measures across all event windows, with the interaction with *Pure Hiring* and *Pure Firing* are mostly insignificant. Overall, the empirical findings reject Hypothesis 3, as the performance of incumbent sub-advisors is unaffected by hiring a new one or firing one.

In Panel C of Table 3.7, we compare the post-hiring performance of the newly-hired sub-advisor relative to the average for all other sub-advisors in multi-managed funds. $NewSub-advisor$ is an indicator variable equal to one if the sub-advisors has been newly hired; and zero otherwise. The coefficients are insignificant across both measures over all event windows, suggesting that post-hiring performance of the new sub-advisor is indifferent from the existing peers. Overall, these results show that changing sub-advisors do not improve a fund's performance.

3.4.4 Impact of Style Drift on Sub-advisor Turnover

Hypothesis 4 states that sub-advisors that deviate from their mandates are penalized by advisors, who care about the style drift of sub-advisors conditional on the returns generated. We test this hypothesis using the same probit framework as in equation (9), except for the inclusion of a measure of sub-advisors' style drift. We examine the impact of sub-advisors' risk-taking deviations relative to its peers on the likelihood of being hired/fired. We estimate the beta loadings from Carhart (1997) four-factor model for all open-ended mutual funds using a rolling 12-month window and compute the monthly average of factor loadings on the size and value factors within each of the Morningstar 3 by 3 categories. The style-drift of a sub-advisor on a given month is captured by *Beta Deviation*, calculated as the square root of the sum of squared deviations of the size and value betas from the sub-advisor's corresponding Morningstar category averages. Sub-advisor and fund Carhart's (1997) alpha up to three years pre-hiring/firing are included to control for the effect of past performance.

Table 3.8 About Here

Our results are reported in Table 3.8. The coefficient on the prior 12-month sub-advisor beta deviation is indistinguishable from zero for hiring decisions in column (1), but positive and significant at the 1% level for firing decisions in column (3). In columns (2) and (4), we add an interaction term between the prior 12-months sub-advisor beta deviation and the fund's factor alpha. The coefficient on this interaction term is negative and significant at the 5% level for hiring cases (column (2)) but not significant for firing cases (column (4)). The hiring results imply that fund families are more likely to appoint a sub-advisor that has been outperforming recently, but the higher is their beta deviation, the less likely they are to be hired.[9] In unreported tests, we decompose the prior 12-month sub-advisor factor alpha into two censored variables denoting above and below-average alpha observations separately. Statistical significance emerges from interaction with only the above-average factor alphas, indicating that the negative and significant coefficient on the interaction term in column (2) mainly comes from the marginal effect on outperforming sub-advisors.

Overall, results suggest that for hiring decisions, fund families tend to select sub-advisors based on past performance and then select those with less risk-taking deviations from their peers. In the case of firing decisions, both persistent underperformance and excessive style drift of sub-advisors are associated with a higher likelihood of being fired. Fund advisors actively monitor sub-advisors and use high-powered incentives to make sure that they stay in line with the stated investment style, consistent with Chen et al. (2013) and Chevalier and Ellison (1999).

[9] Notice that all continuous explanatory variables in Table 3.8 are standardized to have a mean of zero and standard deviation of one, including sub-advisor factor alphas. A positive standardized factor alpha means the performance is above average.

Style Drift Around Sub-advisor Turnover

Contractual externalities make it difficult for principal fund advisors to monitor performance from an outsourced relationship (Chen et al. (2013)). From columns (3)-(4) in Table 3.8, sub-advisors are more likely to be fired for both underperformance and excessive risk-taking. We expect that sub-advisors facing steeper incentives are more concerned with job preservation, so they may alter the risk structure and reduce beta deviation from its peers (Chevalier and Ellison (1999)). In Table 3.9, we examine if sub-advisor turnover affects style drift, using the same multivariate framework employed in Table 3.5, but with *Beta Deviation* as the dependent variable.

Results do not find strong evidence of long-term style drift after sub-advisor changes. Coefficients for *Pure Firing* and *Mixed Hiring&Firing* are not statistically significant across all windows. For *Pure Hiring* events, we find lower beta deviations for 12 and 24 months after turnover, but the effect disappears afterwards. Furthermore, funds with more sub-advisors tend to exhibit less style drift. While competition among sub-advisors does not lead to better performance, it does seem to reduce deviation from the fund's mandates. This is consistent with our earlier findings that fund advisors act as monitors to mitigate style drift.

Table 3.9 About Here

3.5 Conclusion

This paper makes three main contributions. First, we analyse what drives the large number of hiring and firing events and their resulting impact on mutual fund and sub-advisors' returns. Second, we study competition among sub-advisors by exploiting the fact that the existence of sub-advised (or outsourced) funds with two or more sub-advisors effectively introduces competition into the management of these funds. The presence of more than one sub-advisor leads to a split of management fees among sub-advisors, which in turn encourages greater monitoring between them and appears to benefit performance (Moreno et al. (2018)). Exploiting the fact that 32% of the outsourced funds in our sample are managed by two or more sub-advisors, we examine if the performance of the incumbent sub-advisors varies when either one of them is fired or a new sub-advisor is hired and if this performance is different from that of the recently hired/fired sub-advisors. Finally, we study how actively the fund family monitors its sub-advisors and whether deviating from the fund's investment mandate affects a subadvisor's chances of being hired or fired.

We study the effects of managerial turnover and competition on U.S. sub-advised mutual funds. Funds react to poor past performance by trying to change sub-advisors, but they lack the ability to identify outperforming ones. We find that attempts by asset managers to improve the performance of a fund by hiring and firing a sub-advisor are not successful. In fact, similar to retail investors, sub-advisor turnover decisions exhibit return-chasing behaviour, but these changes neither improve subsequent fund returns and risk measures, nor increase future flows into the fund. Outperforming sub-advisors with larger style drift are less likely to be hired, and the more a sub-advisor deviates from its investment mandate, the more likely it is to be fired. Using variation in the number of sub-advisors to study the degree of competition among sub-advisor, we do not find any changes in performance. Sub-advisor turnover does not affect the risk-adjusted returns of incumbent sub-advisors, and their subsequent performance is not different from those of recently hired/fired sub-advisors. Finally, the likelihood that a sub-advisor is fired increases with its degree of style drift and that the likelihood of being hired following outperformance is reduced by the style drift. Rather than outsourcing asset management to improve returns, advisors just ensure that the mandates given to sub-advisors are followed. These results are relevant for investors trying to understand the impact of delegating asset management to sub-advisors and for those studying the effects of changing the number of sub-advisors on a fund.

Figure 3.1. Annualized Sample Means for Four-Factor Alphas around sub-advisor Turnovers

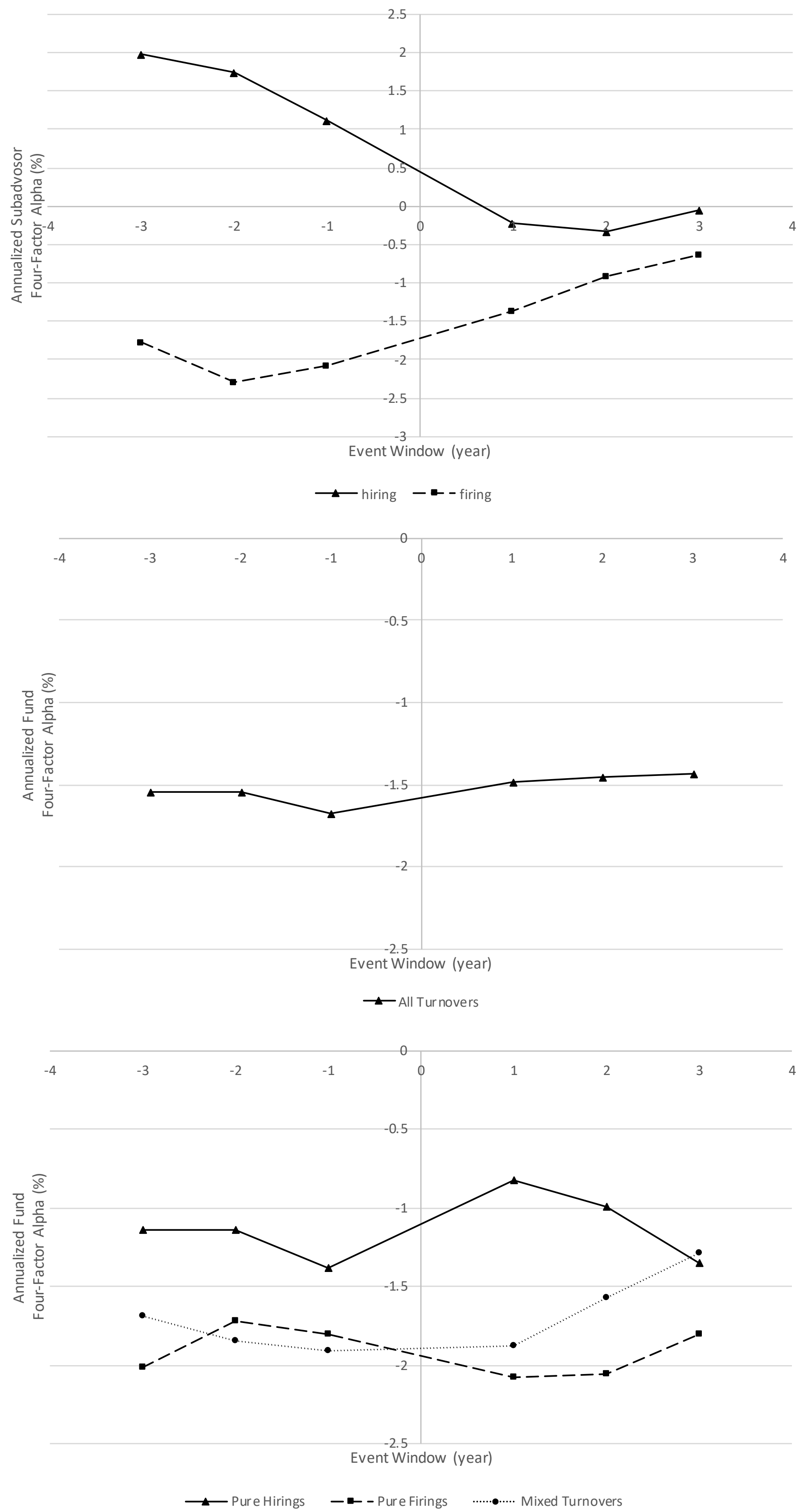
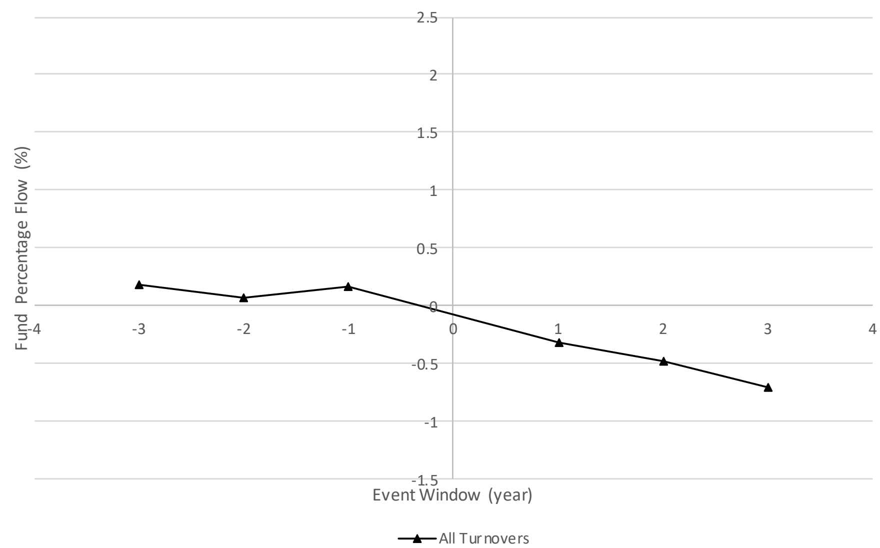


Figure 3.2. Fund Percentage Flows around sub-advisor Turnovers



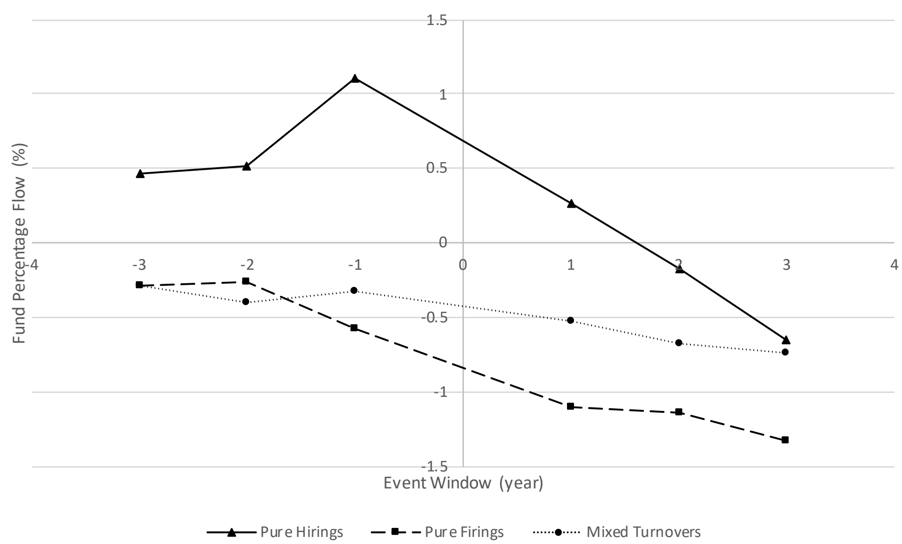


Table 3.1

Descriptive Statistics

This table presents descriptive statistics of U.S. open-ended equity funds and sub-advisors from 1995 to 2016. Panel A contains fund and sub-advisor characteristics. Panel B describes three types of sub-advisor turnover events: "Single/Multi sub-advisors" represents the cases where the fund goes from single-sub-advisor to multi-sub-advisor after hirings, and vice versa for firings; "In-house/Outsource" stands for the case where the fund goes from being managed in-house to being outsourced to a sub-advisor, and vice versa for firing; and "Fund Inception/Delisting" denotes cases where a sub-advisor is hired at inception or fired when the fund is delisted. Panel C decomposes fund turnovers into three mutually exclusive cases, "Pure Hirings/Firings" count the cases where a fund makes only hirings/firings in a given month, "Mixed Hiring&Firing" counts the events where a fund hires and fires at the same time.

Panel A: Fund Advisor and Sub-advisor Characteristics

	Observations	Median	Mean	Std. Dev.	Min	Max
Number of Funds	426					
Number of Sub-advisors	382					
Number of Fund Families	84					
Sub-advisor Carhart's Alpha (%)	249,667	-0.002	0.013	1.833	-5.690	5.943
Sub-advisor Benchmark Alpha (%)	248,407	0.056	0.075	1.895	-5.857	6.197
Fund Carhart's Alpha (%)	251,928	-0.108	-0.111	1.329	-4.371	4.277
Fund Benchmark Alpha (%)	253,140	-0.061	-0.059	1.277	-4.176	4.300
Fund Percentage Flow (%)	198,485	-0.353	0.300	5.145	-13.759	32.004
Fund Flow Alpha ($ billions)	165,573	0.229	1.998	7.149	-13.071	44.788
Fund TNA ($ millions)	246,679	491.8	997.5	1,739.3	0.0	27,786
Number of sub-advisors	334,055	1	2.216	2.624	0	14
Number of Funds Sub-advised by a Sub-advisor	363,563	1	3.253	6.564	0	44
Sub-advisor Beta Deviation	229,907	0.319	0.403	0.316	0.033	1.738
Sub-advisor Tenure (months)	209,602	57	71.199	54.327	3	336

Panel B: Decomposing Sub-advisor Hirings/Firings

	Total Observations	Single/Multi Sub-advisors	In-house/ Outsource	Fund Inception/Delisting
Number of Hirings	1,239	104	105	50
Number of Firings	809	30	229	18

Panel C: Decomposing Fund Turnover

	Total Observations	Pure Hirings	Pure Firings	Mixed Hiring&Firing
Number of Fund Turnovers	1,291	591	374	326

Table 3.2
Sub-advisor Distribution

This table shows the structure of how U.S. open-ended equity funds are managed from 1995 to 2016. For each year, we report the total number of funds, the number of subadvised funds, and the number of those with one, two, three, and four or more sub-advisors. Numbers in brackets show the percentage of sub-advised funds in each category.

Year	All Funds	Sub-advised Funds (%)	Single Sub-advisor (%)	Two Sub-advisors (%)	Three Sub-advisors (%)	Four or More Sub-advisors (%)
1995	1,395	24 (1.72%)	19 (79.17%)	2 (8.33%)	0 (0.00%)	3 (12.50%)
1996	1,568	34 (2.17%)	26 (76.47%)	2 (5.88%)	1 (2.94%)	5 (14.71%)
1997	1,791	39 (2.18%)	30 (76.92%)	2 (5.13%)	0 (0.00%)	7 (17.95%)
1998	2,067	50 (2.42%)	40 (80.00%)	2 (4.00%)	1 (2.00%)	7 (14.00%)
1999	2,285	68 (2.98%)	53 (77.94%)	3 (4.41%)	2 (2.94%)	10 (14.71%)
2000	2,510	93 (3.71%)	69 (74.19%)	9 (9.68%)	3 (3.23%)	12 (12.90%)
2001	2,563	123 (4.80%)	87 (70.73%)	13 (10.57%)	8 (6.50%)	15 (12.20%)
2002	2,503	155 (6.19%)	117 (75.48%)	11 (7.10%)	11 (7.10%)	16 (10.32%)
2003	2,538	173 (6.82%)	126 (72.83%)	15 (8.67%)	13 (7.51%)	19 (10.98%)
2004	2,542	194 (7.63%)	136 (70.10%)	21 (10.82%)	13 (6.70%)	24 (12.37%)
2005	2,659	215 (8.09%)	149 (69.30%)	26 (12.09%)	12 (5.58%)	28 (13.02%)
2006	2,715	230 (8.47%)	157 (68.26%)	27 (11.74%)	19 (8.26%)	27 (11.74%)
2007	2,796	250 (8.94%)	171 (68.40%)	27 (10.80%)	23 (9.20%)	29 (11.60%)
2008	3,021	259 (8.57%)	174 (67.18%)	29 (11.20%)	21 (8.11%)	35 (13.51%)
2009	3,026	268 (8.86%)	179 (66.79%)	30 (11.19%)	19 (7.09%)	40 (14.93%)
2010	2,910	280 (9.62%)	184 (65.71%)	32 (11.43%)	20 (7.14%)	44 (15.71%)
2011	2,948	290 (9.84%)	191 (65.86%)	34 (11.72%)	19 (6.55%)	46 (15.86%)
2012	2,952	307 (10.40%)	202 (65.80%)	39 (12.70%)	18 (5.86%)	48 (15.64%)
2013	2,989	316 (10.57%)	211 (66.77%)	36 (11.39%)	22 (6.96%)	47 (14.87%)
2014	3,000	346 (11.53%)	236 (68.21%)	36 (10.40%)	21 (6.07%)	53 (15.32%)
2015	3,113	350 (11.24%)	233 (66.57%)	37 (10.57%)	22 (6.29%)	58 (16.57%)
2016	3,118	349 (11.19%)	232 (66.48%)	39 (11.17%)	22 (6.30%)	56 (16.05%)
Average	2,591	201 (7.76%)	137 (68.16%)	21 (10.45%)	13 (6.47%)	29 (14.92%)

Table 3.3

The Probability of Sub-advisor Turnover and Past Performance

This table investigates how past performance of a sub-advisor and the performance of the fund affect sub-advisor turnover decisions using panel probit regressions. The dependent variables $Hired_{i,t}$ and $Fired_{i,t}$ are indicator variables that take a value of one if sub-advisor i is, respectively, hired and fired in month t; and zero otherwise. As explanatory variables, we use Carhart's (1997) alphas and the benchmark-adjusted abnormal returns measured in yearly windows in the past 12, 24, and 36 months before each hiring and event. Further control variables include lagged values of *Log(Fund TNA*, *Log(Sub-advisor Age)*, *Log(Fund Age)* and *No. of Funds in Family.* For firing decisions, we also include *Log(Sub-advisor Tenure)* and an indicator variable *Co-branding*, equal to 1 if the sub-advisor's name appears in the fund's name, zero otherwise. All continuous explanatory variables are standardized to have a mean of zero and standard deviation of one. Year fixed effects are included and the standard errors are clustered at the fund-level. ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively.

	$\text{Hired}_{i,t}$		$\text{Fired}_{i,t}$	
Variables	(1)	(2)	(3)	(4)
Sub-advisor Factor Alpha[-12,-1]	0.062***		-0.109***	
	(0.023)		(0.017)	
Sub-advisor Factor Alpha[-24,-13]	0.073***		-0.131***	
	(0.022)		(0.022)	
Sub-advisor Factor Alpha[-36,-25]	0.080***		0.010	
	(0.018)		(0.022)	
Fund Factor Alpha[-12,-1]	-0.082***		-0.073***	
	(0.027)		(0.020)	
Fund Factor Alpha[-24,-13]	-0.038		0.015	
	(0.025)		(0.026)	
Fund Factor Alpha[-36,-25]	-0.039**		-0.028*	
	(0.020)		(0.016)	
Sub-advisor Benchmark Alpha[-12,-1]		0.087***		-0.111***
		(0.017)		(0.017)
Sub-advisor Benchmark Alpha[-24,-13]		0.075***		-0.124***
		(0.021)		(0.020)
Sub-advisor Benchmark Alpha[-36,-25]		0.046**		-0.005
		(0.020)		(0.020)
Fund Benchmark Alpha[-12,-1]		-0.081***		-0.070***
		(0.025)		(0.023)
Fund Benchmark Alpha[-24,-13]		-0.041**		0.044**
		(0.019)		(0.022)
Fund Benchmark Alpha[-36,-25]		-0.027		-0.025*
		(0.016)		(0.015)
$\text{Log(Fund TNA)}_{i,t-1}$	0.020	0.019	0.016	0.016
	(0.017)	(0.017)	(0.010)	(0.011)
$\text{Log(Sub-advisor Age)}_{i,t-1}$	0.009	0.003	-0.004	-0.012*
	(0.010)	(0.010)	(0.006)	(0.006)
$\text{Log(Fund Age)}_{i,t-1}$	-0.034	-0.030	-0.067***	-0.057***
	(0.029)	(0.026)	(0.019)	(0.020)
$\text{No. of Funds in Family}_{i,t-1}$	-0.026*	-0.028**	-0.002	-0.002
	(0.014)	(0.013)	(0.006)	(0.006)
$\text{Log(Sub-advisor Tenure)}_{i,t-1}$			0.005	0.005
			(0.012)	(0.012)
Co-branding			-0.004	-0.019
			(0.022)	(0.023)
Constant	-2.598***	-2.597***	-1.741***	-1.764***
	(0.091)	(0.092)	(0.039)	(0.040)
Observations	127,286	128,974	86,721	87,930
Number of sub-advisors	1,099	1,116	715	723
Year FE	YES		YES	

Table 3.4

Fund Performance, Risk, and Asset Flows around Sub-advisor Turnover Events

This table compares fund return, risk, and asset flows measures. For each dependent measure, we examine event-windows for 12, 24, and 36 months before and after sub-advisor turnover. Panel A examines fund return measures: Carhart's (1997) alpha, benchmark-adjusted abnormal returns, and Sharpe ratios. Panel B compares fund return volatilities and two flow measures. *Percentage Flow* is the dollar flow scaled by the start-of-period fund total net assets (TNA). *Flow Alpha* is a measure of abnormal flow defined as the difference between realized flow and expected flow predicted from a regression based on lagged returns, fund age, and fund TNA. *Pure Hiring* is an indicator variable equal to one if a fund hires a sub-advisor in a given month without firing any sub-advisor; and zero otherwise. *Pure Firing* is an indicator variable equal to one if a fund hires a sub-advisor in a given month without firing any sub-advisor; and zero otherwise. Only funds with continuous return data available for all the months in an event window are included. Hiring events that signal a fund going from being in-house managed to being outsourced are excluded from the sample, so do the firing events of an outsourced fund that makes the fund to be managed in-house. *Post Turnover* is an indicator variable which takes a value of one if the sample mean is taken after the turnover event; zero otherwise. Heteroskedasticity-robust standard errors are reported in parentheses. ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively.

Panel A: Fund Return Measures

	Fund Factor Alpha			Fund Benchmark Alpha			Fund Sharpe Ratio		
Variables	[-12,12]	[-24,24]	[-36,36]	[-12,12]	[-24,24]	[-36,36]	[-12,12]	[-24,24]	[-36,36]
Post Turnover	0.003	0.023	0.033*	0.034	0.021	0.007	0.023	0.067**	0.123***
	(0.031)	(0.024)	(0.019)	(0.036)	(0.028)	(0.023)	(0.036)	(0.028)	(0.023)
Post Turnover*Pure Hiring	0.043	-0.010	-0.051	0.001	-0.046	-0.055	-0.028	-0.057	-0.056*
	(0.046)	(0.034)	(0.033)	(0.055)	(0.041)	(0.037)	(0.051)	(0.040)	(0.032)
Post Turnover*Pure Firing	-0.026	-0.050	-0.016	0.007	-0.011	0.012	0.042	0.027	-0.016
	(0.055)	(0.042)	(0.037)	(0.061)	(0.050)	(0.043)	(0.065)	(0.047)	(0.038)
Pure Hiring	0.045	0.058**	0.045*	0.054	0.091***	0.077***	-0.010	0.030	0.032
	(0.033)	(0.026)	(0.025)	(0.040)	(0.031)	(0.028)	(0.037)	(0.029)	(0.022)
Pure Firing	0.009	0.011	-0.027	-0.023	-0.013	-0.034	0.021	0.018	0.019
	(0.041)	(0.031)	(0.027)	(0.047)	(0.040)	(0.033)	(0.048)	(0.036)	(0.027)
Constant	-0.159***	-0.154***	-0.140***	-0.116***	-0.096***	-0.076***	0.177***	0.114***	0.055***
	(0.022)	(0.018)	(0.013)	(0.027)	(0.022)	(0.017)	(0.027)	(0.021)	(0.016)
Observations	936	798	640	940	802	654	956	818	676
R^2	0.014	0.020	0.017	0.010	0.022	0.026	0.007	0.016	0.071
Balanced Sample	YES	YES	YES	YES	YES	YES	YES	YES	YES

(Continued)

Table 3.4 - Continued

Panel B: Fund Risk and Flow Measures

	Fund Return Volatility			Fund Percentage Flow			Fund Flow Alpha		
	[-12,12]	[-24,24]	[-36,36]	[-12,12]	[-24,24]	[-36,36]	[-12,12]	[-24,24]	[-36,36]
Post Turnover	-0.147	-0.394**	-0.553***	-0.255	-0.288	-0.551**	2.199	0.991	-0.005
	(0.213)	(0.196)	(0.189)	(0.270)	(0.264)	(0.253)	(1.868)	(1.542)	(1.945)
Post Turnover*Pure Hiring	0.157	0.111	0.048	-0.578	-0.377	-0.540	-0.139	0.055	0.448
	(0.292)	(0.266)	(0.261)	(0.490)	(0.385)	(0.392)	(2.318)	(1.653)	(2.087)
Post Turnover*Pure Firing	-0.250	-0.200	-0.125	-0.050	-0.496	-0.565	0.197	0.677	1.184
	(0.353)	(0.315)	(0.299)	(0.392)	(0.382)	(0.404)	(2.689)	-1.985	(2.325)
Pure Hiring	0.324	0.354*	0.341*	1.463***	0.813***	0.727**	-0.810	-0.972	-1.688
	(0.207)	(0.187)	(0.180)	(0.417)	(0.308)	(0.321)	(1.179)	(1.153)	(1.742)
Pure Firing	0.181	0.155	0.320	-0.562*	-0.062	-0.082	-0.532	-0.888	-1.793
	(0.266)	(0.232)	(0.214)	(0.305)	(0.317)	(0.348)	(1.283)	(1.157)	(1.745)
Constant	4.713***	4.966***	5.159***	-0.255	-0.208	-0.058	1.079	1.332	2.103
	(0.153)	(0.142)	(0.133)	(0.206)	(0.210)	(0.200)	(1.051)	(1.109)	(1.708)
Observations	826	732	618	688	560	448	74	56	40
R^2	0.013	0.031	0.051	0.079	0.059	0.093	0.063	0.056	0.067
Balanced Sample	YES	YES	YES	YES	YES	YES	YES	YES	YES

Table 3.5

Effect of Sub-advisor Turnover on Fund Performance

This table examines the effect of sub-advisor turnover on future fund performance. The dependent variables are the Carhart's (1997) alphas, the fund's benchmark alpha, and the Sharpe ratio measured over the next 12, 24, and 36 months. We distinguish the differential turnover effects for three distinct types of sub-advisor turnover, indicated by three binary variables *Pure Hiring*, *Pure Firing* and *Mixed Hiring/Firing*, respectively. *Pure Hiring* is an indicator variable equal to one if a fund hires a sub-advisor in a given month without firing any sub-advisor; and zero otherwise. *Pure Firing* is an indicator variable equal to one if a fund hires a sub-advisor in a given month without firing any sub-advisor; and zero otherwise. Fund and year fixed effects are included and the standard errors are clustered at the fund-level. ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively.

	Fund Factor Alpha			Fund Benchmark Alpha			Fund Sharpe Ratio		
Variables	[1,12]	[1,24]	[1,36]	[1,12]	[1,24]	[1,36]	[1,12]	[1,24]	[1,36]
Pure Hiring	0.022	0.014	0.009	0.032	0.012	0.013	-0.165	-0.151	-0.145
	(0.025)	(0.018)	(0.016)	(0.026)	(0.014)	(0.013)	(0.200)	(0.196)	(0.197)
Pure Firing	-0.059**	-0.024	-0.015	-0.062***	-0.026	-0.021	-0.234	-0.237	-0.256
	(0.027)	(0.019)	(0.017)	(0.024)	(0.020)	(0.020)	(0.262)	(0.265)	(0.266)
Mixed Hiring/Firing	-0.027	-0.022	-0.016	-0.024	-0.032**	-0.025**	-0.089	-0.096	-0.087
	(0.019)	(0.014)	(0.011)	(0.019)	(0.013)	(0.010)	(0.105)	(0.105)	(0.105)
Fund Factor Alpha[-12,-1]	-0.079***	-0.073***	-0.086***						
	(0.018)	(0.013)	(0.011)						
Fund Factor Alpha[-24,-13]	-0.061***	-0.078***	-0.067***						
	(0.015)	(0.013)	(0.012)						
Fund Factor Alpha[-36,-25]	-0.051***	-0.038***	-0.034***						
	(0.014)	(0.012)	(0.010)						
Fund Benchmark Alpha[-12,-1]				-0.107***	-0.080***	-0.062***			
				(0.018)	(0.015)	(0.014)			
Fund Benchmark Alpha[-24,-13]				-0.042**	-0.023	-0.026*			
				(0.019)	(0.015)	(0.014)			
Fund Benchmark Alpha[-36,-25]				0.005	-0.009	-0.021*			
				(0.017)	(0.013)	(0.011)			
Fund Sharpe Ratio[-12,-1]							-0.223***	-0.154*	-0.060
							(0.085)	(0.085)	(0.085)
Fund Sharpe Ratio[-24,-13]							-0.180*	-0.078	0.011
							(0.099)	(0.098)	(0.098)
Fund Sharpe Ratio[-36,-25]							0.010	0.053	0.030
							(0.046)	(0.049)	(0.047)
Log(Fund TNA)$_{i,t-1}$	-0.071***	-0.064***	-0.059***	-0.068***	-0.058***	-0.056***	-0.154**	-0.139**	-0.132*
	(0.013)	(0.012)	(0.011)	(0.014)	(0.013)	(0.012)	(0.067)	(0.067)	(0.067)
% of Fund's TNA relative to Fund Family's TNA $_{i,t-1}$	-0.156	-0.081	-0.024	-0.096	-0.035	0.040	0.109	0.067	0.056
	(0.139)	(0.124)	(0.113)	(0.173)	(0.157)	(0.145)	(0.192)	(0.188)	(0.186)
Log(Fund Age)$_{i,t-1}$	-0.098***	-0.083***	-0.078***	-0.016	-0.032	-0.037	0.081	0.073	0.077
	(0.033)	(0.030)	(0.029)	(0.038)	(0.033)	(0.030)	(0.092)	(0.092)	(0.091)
No.(Sub-advisors in Fund)$_{i,t-1}$	0.008	0.005	0.005	0.009	0.007	0.005	0.012	0.016	0.016
	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)	(0.011)	(0.014)	(0.014)	(0.014)
Constant	1.211***	1.241***	1.309***	1.104***	1.030***	1.122***	3.131***	2.722**	2.495**
	(0.269)	(0.256)	(0.253)	(0.257)	(0.241)	(0.235)	(1.089)	(1.088)	(1.090)
Observations	36,143	36,143	36,143	36,587	36,587	36,587	36,880	36,880	36,880
Number of Funds	295	295	295	300	300	300	302	302	302
R^2	0.104	0.150	0.175	0.083	0.127	0.156	0.001	0.001	0.000
Fund FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Table 3.6

Effect of Sub-advisor Turnover on Fund Volatility and Flows

This table examines the turnover effect on future fund return volatility and flows. The dependent variables are fund return volatility, percentage flow, and flow alpha, measured over the next 12, 24, and 36 months. We distinguish the differential turnover effects on flows for three distinct types of sub-advisor turnover, indicated by three binary variables *Pure Hiring*, *Pure Firing* and *Mixed Hiring/Firing*, respectively. *Pure Hiring* is an indicator variable equal to one if a fund hires a sub-advisor in a given month without firing any sub-advisor; and zero otherwise. *Pure Firing* is an indicator variable equal to one if a fund hires a sub-advisor in a given month without hiring any other sub-advisor; and zero otherwise. Lagged values of the dependent variable in each of the three years before the event are included as controls. Fund and year fixed effects are included and the standard errors are clustered at the fund-level. ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively.

	Fund Return Volatility			Fund Percentage Flow			Fund Flow Alpha		
Variables	[1,12]	[1,24]	[1,36]	[1,12]	[1,24]	[1,36]	[1,12]	[1,24]	[1,36]
Pure Hiring	0.028	-0.013	-0.027	0.413**	0.219*	0.046	0.389	0.521	0.328
	(0.054)	(0.048)	(0.040)	(0.166)	(0.118)	(0.092)	(0.606)	(0.361)	(0.374)
Pure Firing	-0.108*	-0.054	-0.024	-0.168	-0.100	-0.111	0.401	0.344	0.315
	(0.063)	(0.048)	(0.039)	(0.148)	(0.141)	(0.120)	(0.566)	(0.402)	(0.330)
Mixed Hiring/Firing	0.083*	0.039	0.023	0.139	0.134	0.089	-0.299	-0.372**	-0.204
	(0.042)	(0.029)	(0.021)	(0.104)	(0.083)	(0.067)	(0.232)	(0.166)	(0.177)
Fund Return Volatility[-12,-1]	0.051***	-0.033**	-0.031**						
	(0.015)	(0.013)	(0.012)						
Fund Return Volatility[-24,-13]	-0.059***	-0.036***	-0.039***						
	(0.011)	(0.010)	(0.010)						
Fund Return Volatility[-36,-25]	-0.010	-0.022**	-0.038***						
	(0.010)	(0.010)	(0.010)						
Fund Percentage Flow[-12,-1]				0.275***	0.192***	0.134***			
				(0.032)	(0.026)	(0.023)			
Fund Percentage Flow[-24,-13]				0.052***	0.030**	0.019*			
				(0.019)	(0.014)	(0.010)			
Fund Percentage Flow[-36,-25]				0.010	0.007	0.002			
				(0.012)	(0.010)	(0.008)			
Fund Flow Alpha[-12,-1]							0.052	0.066	0.038
							(0.058)	(0.043)	(0.041)
Fund Flow Alpha[-24,-13]							0.124**	0.074	0.045
							(0.055)	(0.045)	(0.049)
Fund Flow Alpha[-36,-25]							0.046	0.020	-0.001
							(0.070)	(0.071)	(0.061)
Fund Factor Alpha[-12,-1]	0.093**	0.064**	0.061**	0.845***	0.680***	0.578***	-0.008	-0.006	-0.009
	(0.037)	(0.030)	(0.024)	(0.106)	(0.093)	(0.080)	(0.166)	(0.164)	(0.167)
Fund Factor Alpha[-24,-13]	0.053*	0.062**	0.060**	0.372***	0.331***	0.264***	-0.184	-0.140	-0.048
	(0.030)	(0.026)	(0.024)	(0.098)	(0.076)	(0.064)	(0.189)	(0.192)	(0.175)
Fund Factor Alpha[-36,-25]	0.045**	0.039**	0.021	0.124	0.124*	0.095*	-0.171	-0.055	-0.027
	(0.021)	(0.019)	(0.018)	(0.078)	(0.065)	(0.050)	(0.190)	(0.164)	(0.164)
Log(Fund TNA)$_{i,t-1}$	0.116***	0.089***	0.067**	-1.135***	-1.138***	-1.083***	0.967***	1.053***	1.058***
	(0.032)	(0.032)	(0.030)	(0.111)	(0.099)	(0.091)	(0.212)	(0.235)	(0.240)
% of Fund's TNA relative to Fund Family's TNA $_{i,t-1}$	-0.381	-0.245	-0.154	-1.149	-1.259	-1.213	3.262**	3.074*	2.812*
	(0.332)	(0.336)	(0.307)	(0.973)	(0.999)	(0.951)	(1.636)	(1.688)	(1.684)
Log(Fund Age)$_{i,t-1}$	0.013	0.027	0.031	-0.289	-0.367	-0.391*	-2.564***	-2.704**	-2.992**
	(0.070)	(0.074)	(0.072)	(0.249)	(0.246)	(0.235)	(0.980)	(1.095)	(1.188)
No.(Sub-advisors in Fund)$_{i,t-1}$	-0.027	-0.034	-0.026	0.013	-0.048	-0.065	0.345	0.240	0.143
	(0.021)	(0.023)	(0.023)	(0.071)	(0.065)	(0.055)	(0.388)	(0.299)	(0.258)
Constant	3.860***	4.447***	5.329***	22.332***	21.925***	20.988***	-17.771***	-19.312***	-18.486***
	(0.628)	(0.610)	(0.552)	(1.937)	(1.753)	(1.628)	(4.488)	(5.047)	(5.094)
Observations	36,016	36,016	36,016	33,159	33,159	33,159	30,000	30,000	30,000
Number of Funds	293	293	293	0.289	0.331	0.345	0.097	0.129	0.135
R^2	0.766	0.834	0.860	290	290	290	284	284	284
Fund FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Table 3.7

Effect of Sub-advisor Turnover on Sub-advisors' Performance

This table compares the performance of sub-advisors around sub-advisor turnover events using 12, 24, and 36 months event-windows. Panel A examines the performance of sub-advisors that have been just hired and fired, Panel B explores the performance of other sub-advisors in multi-advised funds when a fund hires/fires a sub-advisor, and Panel C compares the performance of a newly-hired sub-advisor with the fund's other existing sub-advisors. *Pure Hiring* is an indicator variable equal to one if a fund hires a sub-advisor in a given month without firing any sub-advisor; and zero otherwise. *Pure Firing* is an indicator variable equal to one if a fund hires a sub-advisor in a given month without firing any sub-advisor; and zero otherwise. In Panel A, *Single_Multi* indicates a fund going from having a single sub-advisor to multiple sub-advisors after hiring and vice versa after firing events. In Panel B, *Post Turnover* is an indicator variable equal to one if the observation is after the turnover event; zero otherwise. In Panel C, *New Sub-advisor* is an indicator variable taking a value of one if the sample mean is taken for the newly-hired sub-advisor; and zero otherwise. Only funds with continuous return data available for all the months in an event window are included. Events with sub-advisors that are fired within 12, 24, and 36 months post-hiring are excluded. Heteroskedasticity-robust standard errors are reported in parentheses. ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively.

Panel A: Sub-advisor Performance after Turnover

	Sub-advisor Factor Alpha						Sub-advisor Benchmark Alpha					
	Hiring Events			Firing Events			Hiring Events			Firing Events		
Variable	[-12,12]	[-24,24]	[-36,36]	[-12,12]	[-24,24]	[-36,36]	[-12,12]	[-24,24]	[-36,36]	[-12,12]	[-24,24]	[-36,36]
Post Hiring	-0.096***	-0.167***	-0.169***				-0.127***	-0.190***	-0.183***			
	(0.027)	(0.025)	(0.024)				(0.031)	(0.026)	(0.025)			
Post Firing				0.067*	0.127***	0.101**				0.105**	0.151***	0.077**
				(0.040)	(0.035)	(0.039)				(0.044)	(0.035)	(0.039)
Post Hiring/Firing* Single_Multi	-0.184**	-0.060	-0.006	-0.119	-0.205**	-0.099	-0.219*	-0.002	0.034	-0.188*	-0.200**	-0.119
	(0.086)	(0.083)	(0.080)	(0.113)	(0.097)	(0.088)	(0.115)	(0.095)	(0.089)	(0.111)	(0.096)	(0.096)
Single_Multi	0.097	0.043	-0.009	0.064	0.201***	0.076	0.108	0.008	-0.043	0.049	0.141**	0.074
	(0.059)	(0.058)	(0.057)	(0.075)	(0.058)	(0.064)	(0.089)	(0.076)	(0.069)	(0.090)	(0.067)	(0.062)
Constant	0.085***	0.141***	0.166***	-0.177***	-0.203***	-0.152***	0.168***	0.219***	0.249***	-0.112***	-0.153***	-0.091***
	(0.020)	(0.019)	(0.019)	(0.028)	(0.026)	(0.032)	(0.023)	(0.019)	(0.020)	(0.033)	(0.027)	(0.030)
Observations	1,714	1,286	972	716	508	284	1,712	1,292	974	712	508	284
R^2	0.014	0.041	0.055	0.004	0.031	0.024	0.016	0.043	0.056	0.009	0.038	0.015
Balanced Sample	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

cmidrule(lr)14-16

(Continued)

Table 3.7 - Continued

Panel B: Existing/Remaining Sub-advisors' Performance after Turnover

	Sub-advisor Factor Alpha			Sub-advisor Benchmark Alpha		
Variable	[-12,12]	[-24,24]	[-36,36]	[-12,12]	[-24,24]	[-36,36]
Pure Hiring	0.063***	0.093***	0.033	0.081***	0.095***	0.052*
	(0.023)	(0.024)	(0.027)	(0.027)	(0.026)	(0.028)
Pure Firing	-0.023	0.021	-0.043	-0.048	-0.013	-0.033
	(0.027)	(0.027)	(0.030)	(0.030)	(0.031)	(0.035)
Post Turnover	-0.056**	-0.020	-0.048**	-0.023	-0.023	-0.051*
	(0.022)	(0.022)	(0.025)	(0.026)	(0.025)	(0.029)
Post Turnover*Pure Hiring	0.013	-0.041	0.020	-0.002	-0.035	0.007
	(0.032)	(0.032)	(0.036)	(0.037)	(0.036)	(0.039)
Post Turnover*Pure Firing	0.034	-0.018	0.083**	0.038	0.017	0.061
	(0.036)	(0.035)	(0.040)	(0.040)	(0.040)	(0.045)
Constant	-0.020	-0.018	0.040**	0.057***	0.078***	0.120***
	(0.016)	(0.016)	(0.018)	(0.019)	(0.019)	(0.022)
Observations	6,500	3,576	2,056	6,496	3,586	2,058
R^2	0.005	0.009	0.006	0.005	0.009	0.008
Balanced Sample	YES	YES	YES	YES	YES	YES

Panel C: Comparing Post-Hiring Performance between New and Existing Sub-advisors

	Sub-advisor Factor Alpha			sub-advisor Benchmark Alpha		
New Sub-advisor	0.055	0.001	0.001	-0.006	-0.044	-0.010
	(0.036)	(0.031)	(0.030)	(0.042)	(0.036)	(0.033)
Constant	-0.016	0.016	0.029	0.070**	0.112***	0.114***
	(0.026)	(0.022)	(0.024)	(0.032)	(0.029)	(0.025)
Observations	731	669	585	729	668	587
R^2	0.002	0.000	0.000	0.000	0.002	0.000
Balanced Sample	YES	YES	YES	YES	YES	YES

Table 3.8

Probability of Sub-advisor Turnover and Style drift

This table investigates the effect of sub-advisor style drift on hiring/firing decisions. $Hired_{i,t}$ and $Fired_{i,t}$ are indicator variables that take a value of one if sub-advisor i is, respectively, hired and fired in month t; and zero otherwise. As explanatory variables, we measure the level of style drift for a sub-advisor, *Sub-advisor Beta Deviation*, by the squared root of beta deviations on the size and value dimensions, where the beta loadings are estimated using a 12-month rolling window. We included fund and sub-advisor Carhart's (1997) alphas for each for the past three years. Further control variables include lagged values of *Log(Fund TNA*, *Log(Sub-advisor Age)*, *Log(Fund Age)* and *No. of Funds in Family.* For firing decisions, we also include *Log(Sub-advisor Tenure)* and an indicator variable *Co-branding*, equal to 1 if the sub-advisor's name appears in the fund's name, zero otherwise. All continuous explanatory variables are standardized to have a mean of zero and standard deviation of one. Year fixed effects are included and the standard errors are clustered at the fund-level. ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively.

	$\text{Hired}_{i,t}$		$\text{Fired}_{i,t}$	
Variables	(1)	(2)	(3)	(4)
Sub-advisor Beta Deviation[-12,-1]* Sub-advisor Factor Alpha[-12,-1]		-0.028**		0.004
		(0.012)		(0.011)
Sub-advisor Beta Deviation [-12,-1]	0.022	0.024	0.044***	0.045***
	(0.018)	(0.017)	(0.015)	(0.015)
Sub-advisor Factor Alpha[-12,-1]	0.050**	0.065***	-0.113***	-0.115***
	(0.023)	(0.024)	(0.016)	(0.018)
Sub-advisor Factor Alpha[-24,-13]	0.054**	0.055***	-0.113***	-0.112***
	(0.021)	(0.021)	(0.020)	(0.020)
Sub-advisor Factor Alpha[-36,-25]	0.081***	0.080***	-0.003	-0.003
	(0.017)	(0.017)	(0.021)	(0.021)
Fund Factor Alpha[-12,-1]	-0.075***	-0.079***	-0.065***	-0.064***
	(0.026)	(0.026)	(0.019)	(0.020)
Fund Factor Alpha[-24,-13]	-0.033	-0.033	0.015	0.015
	(0.023)	(0.023)	(0.024)	(0.024)
Fund Factor Alpha[-36,-25]	-0.041**	-0.042**	-0.027	-0.027
	(0.021)	(0.020)	(0.018)	(0.018)
Log(Fund TNA)$_{i,t-1}$	0.011	0.012	0.021*	0.021*
	(0.018)	(0.018)	(0.011)	(0.011)
Log(Sub-advisor Age)$_{i,t-1}$	0.011	0.011	-0.005	-0.004
	(0.011)	(0.010)	(0.007)	(0.007)
Log(Fund Age)$_{i,t-1}$	-0.034	-0.034	-0.075***	-0.075***
	(0.030)	(0.030)	(0.019)	(0.019)
Number of Funds in the Family$_{i,t-1}$	-0.026*	-0.026*	-0.001	-0.001
	(0.014)	(0.014)	(0.006)	(0.006)
Log(Sub-advisor Tenure)$_{i,t-1}$			0.002	0.002
			(0.012)	(0.012)
Co-branding			0.008	0.008
			(0.025)	(0.025)
Constant	-2.587***	-2.586***	-1.732***	-1.732***
	(0.096)	(0.096)	(0.040)	(0.040)
Observations	117,698	117,698	80,710	80,710
Number of sub-advisors	1,017	1,017	664	664
Wald χ^2	177.2	185.7	772	773.8
Year FE	YES		YES	

Table 3.9

Effect of Sub-advisor Turnover on Style Drift

This table compares the style drift of a fund around sub-advisor turnover events using 12, 24, and 36 months event-windows. *Beta Deviation* is calculated as the the squared root of beta deviations on the size and value dimensions, where the beta loadings on thensize and value factors are estimated using a 12-month rolling window. *Pure Hiring* is an indicator variable equal to one if a fund hires a sub-advisor in a given month without firing any sub-advisor; and zero otherwise. *Pure Firing* is an indicator variable equal to one if a fund hires a sub-advisor in a given month without firing any sub-advisor; and zero otherwise. Hiring events that signal a fund going from in-house managed to outsourced are excluded from the sample, so do the firing events marking an outsourced funds getting back to in-house management. Only funds with continuous return data available for all the months in an event window are included. Fund and year fixed effects are included and the standard errors are clustered at the fund-level. ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively.

	Fund Beta Deviation		
Variables	[1,12]	[1,24]	[1,36]
Pure Hiring	-0.023**	-0.011*	-0.007
	(0.011)	(0.007)	(0.006)
Pure Firing	-0.003	0.001	0.003
	(0.011)	(0.007)	(0.006)
Mixed Hiring&Firing	0.002	-0.003	0.003
	(0.010)	(0.006)	(0.005)
Fund Beta Deviation [-12,-1]	0.041**		
	(0.016)		
Fund Beta Deviation [-24,-13]		0.003	
		(0.031)	
Fund Beta Deviation [-36,-25]			-0.037
			(0.034)
Fund Factor Alpha[-12,-1]	0.005	0.009	0.001
	(0.008)	(0.006)	(0.006)
Fund Factor Alpha[-24,-13]	-0.004	-0.006	-0.009*
	(0.008)	(0.006)	(0.005)
Fund Factor Alpha[-36,-25]	-0.004	-0.003	-0.004
	(0.004)	(0.004)	(0.004)
Log(Fund TNA)$_{i,t-1}$	-0.003	-0.000	-0.000
	(0.005)	(0.005)	(0.005)
% of Fund's TNA relative to Fund Family's TNA $_{i,t-1}$	-0.017	0.033	0.031
	(0.039)	(0.037)	(0.038)
Log(Fund Age)$_{i,t-1}$	0.006	-0.009	-0.013
	(0.013)	(0.013)	(0.013)
No.(Sub-advisors in Fund)$_{i,t-1}$	-0.011***	-0.007*	-0.005
	(0.004)	(0.004)	(0.004)
Constant	0.377***	0.322***	0.313***
	(0.103)	(0.100)	(0.107)
Observations	34,883	34,883	34,883
Number of Funds	293	293	293
R^2	0.128	0.172	0.214
Fund FE	YES	YES	YES
Time FE	YES	YES	YES

Table A3.1
3x3 Morningstar Categories and Benchmarks

	Value	Blend	Growth
Large	Russell 1000 Value TR USD	Russell 1000 TR USD	Russell 1000 Growth TR USD
MidCap	Russell Mid Cap Value TR USD	Russell Mid Cap TR USD	Russell Mid Cap Growth TR USD
Small	S&P SmallCap 600 Value TR USD	Russell 2000 TR USD	S&P SmallCap 600 Growth TR USD

Table A3.2
Largest sub-advisors

This table lists names of the top 20 sub-advisors sorted in descending order by the number of funds that they sub-advise for. The third column shows the total number of family complexes these top funds sub-advise for whilst the last column shows assets under management of the sub-advisory firms at the time when they were sub-contracted. These top 20 sub-advisers cover approximately 30% of all the sub-advisory activity in our sample.

Rank	Sub-advisor Name	Funds it sub-advises for	Families it sub-advises for	Sub-advisor's family AUM
1.	Wellington Management Company, LP	59	21	444,673.4
2.	Blackrock Investment Management, LLC	39	12	1,793,967.8
3.	Alliance Bernstein, LP	30	12	428,378.4
4.	T. Rowe Price Associates, Inc.	29	14	256,298.3
5.	Goldman Sachs Asset Management, LP	21	13	559,907.3
6.	J.P. Morgan Investment Management, Inc.	21	12	977,715.8
7.	SSgA Funds Management, Inc.	21	12	263,414.3
8.	Barrow Hanley Mewhinney & Strauss, LLC	19	12	54,948.3
9.	Loomis Sayles & Company, LP	18	13	135,946.6
10.	Mellon Capital Management Corporation	18	7	185,782.1
11.	LSV Asset Management	17	8	37,362.1
12.	TCW Investment Management, Co.	17	13	80,817.5
13.	AJO, LP	15	6	12,626.8
14.	Wells Capital Management, Inc.	15	7	251,030.4
15.	Marsico Capital Management, LLC	14	11	37,684.1
16.	Invesco Advisers, Inc.	13	10	334,794.8
17.	Morgan Stanley Investment, Inc.	13	9	304977.9
18.	Brandywine Global Investment Management, LLC	12	6	41,430.1
19.	Columbus Circle Investors	12	8	11660.6
20.	Systematic Financial Management, LP	12	9	8097.2

Table A3.3

Cross-sectional Variation in Sub-advisor Turnover Effects

This table examines the cross-sectional variation in turnover effects on future fund performance and flows under a panel regression framework. In Panel A, the dependent variables are Carhart's (1997) four-factor abnormal returns (*Fund Factor Alpha*) measured over the next 12, 24, and 36 months. In Panel B, the dependent variables are fund flow alphas measured over the next 12, 24, and 36 months. The corresponding measures for the past 12, 24, and 36 months are included in the explanatory variables to account for any effect of mean reversion or momentum. Interaction terms with various fund characteristics are included to investigate the cross-sectional difference in turnover effects. Fund and year fixed effects are included and the standard errors are clustered at the fund-level. ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively.

Panel A: Size and Importance of Fund for the Fund Family

	Fund Factor Alpha					
Variables	[1,12]	[1,24]	[1,36]	[1,12]	[1,24]	[1,36]
Turnover	-0.477**	-0.340**	-0.228*	-0.033**	-0.015	-0.013
	(0.203)	(0.156)	(0.121)	(0.016)	(0.013)	(0.011)
Turnover*Log(Fund TNA)$_{i,t-1}$	0.023**	0.016**	0.011*			
	(0.010)	(0.008)	(0.006)			
Turnover*Fund's % of Family TNA$_{i,t-1}$				0.103*	0.024	0.031
				(0.060)	(0.039)	(0.038)
Fund Factor Alpha[-12,-1]	-0.079***	-0.073***	-0.086***	-0.079***	-0.073***	-0.086***
	(0.018)	(0.013)	(0.011)	(0.018)	(0.013)	(0.011)
Fund Factor Alpha[-24,-13]	-0.060***	-0.078***	-0.067***	-0.060***	-0.078***	-0.067***
	(0.015)	(0.013)	(0.012)	(0.015)	(0.013)	(0.012)
Fund Factor Alpha[-36,-25]	-0.051***	-0.038***	-0.034***	-0.051***	-0.038***	-0.034***
	(0.014)	(0.012)	(0.010)	(0.014)	(0.012)	(0.010)
Log(Fund TNA)$_{i,t-1}$	-0.071***	-0.064***	-0.059***	-0.071***	-0.064***	-0.059***
	(0.013)	(0.012)	(0.011)	(0.013)	(0.012)	(0.011)
Fund's % of Family TNA$_{i,t-1}$	-0.157	-0.081	-0.024	-0.157	-0.081	-0.024
	(0.139)	(0.124)	(0.113)	(0.139)	(0.124)	(0.113)
Log(Fund Age)$\_i, t-1$	-0.098***	-0.083***	-0.078***	-0.098***	-0.083***	-0.078***
	(0.033)	(0.030)	(0.029)	(0.033)	(0.030)	(0.029)
No.(Sub-advisors in Fund)$_{i,t-1}$	0.008	0.005	0.005	0.007	0.005	0.005
	(0.013)	(0.013)	(0.011)	(0.013)	(0.013)	(0.011)
Constant	1.215***	1.244***	1.311***	1.211***	1.241***	1.309***
	(0.269)	(0.257)	(0.253)	(0.269)	(0.257)	(0.253)
Observations	36,143	36,143	36,143	36,143	36,143	36,143
Number of Funds	295	295	295	295	295	295
R^2	0.104	0.150	0.175	0.103	0.150	0.175
Fund FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES

(Continued)

Table A3.3 - Continued

Panel B: Fund Performance and Flow

	Fund Flow Alpha					
Variables	[1,12]	[1,24]	[1,36]	[1,12]	[1,24]	[1,36]
Turnover	-0.006	0.038	0.067	0.396*	0.236	0.257*
	(0.188)	(0.149)	(0.153)	(0.214)	(0.163)	(0.139)
Turnover*Fund Factor Alpha[-12,-1]	-0.404	-0.056	0.013			
	(0.449)	(0.371)	(0.330)			
Turnover*Fund Flow Alpha[-12,-1]				-0.169**	-0.094	-0.095*
				(0.078)	(0.077)	(0.056)
Fund Flow Alpha[-12,-1]	0.052	0.066	0.038	0.055	0.068	0.040
	(0.058)	(0.043)	(0.041)	(0.058)	(0.043)	(0.041)
Fund Flow Alpha[-24,-13]	0.124**	0.074	0.045	0.123**	0.073	0.045
	(0.055)	(0.045)	(0.049)	(0.055)	(0.045)	(0.049)
Fund Flow Alpha[-36,-25]	0.046	0.020	-0.001	0.046	0.020	-0.002
	(0.070)	(0.071)	(0.061)	(0.070)	(0.071)	(0.061)
Fund Factor Alpha[-12,-1]	-0.004	-0.005	-0.008	-0.005	-0.004	-0.007
	(0.166)	(0.164)	(0.168)	(0.166)	(0.163)	(0.167)
Fund Factor Alpha[-24,-13]	-0.183	-0.138	-0.046	-0.182	-0.137	-0.046
	(0.189)	(0.192)	(0.175)	(0.189)	(0.192)	(0.175)
Fund Factor Alpha[-36,-25]	-0.171	-0.055	-0.027	-0.173	-0.056	-0.028
	(0.190)	(0.164)	(0.165)	(0.190)	(0.164)	(0.164)
Log(Fund TNA)$_{i,t-1}$	0.968***	1.055***	1.059***	0.967***	1.055***	1.058***
	(0.212)	(0.235)	(0.240)	(0.212)	(0.235)	(0.240)
% of Fund's TNA relative to Fund Family's TNA $_{i,t-1}$	3.267**	3.078*	2.814*	3.274**	3.083*	2.820*
	(1.636)	(1.688)	(1.684)	(1.636)	(1.688)	(1.685)
Log(Fund Age)$_{i,t-1}$	-2.560***	-2.701**	-2.990**	-2.565***	-2.703**	-2.992**
	(0.981)	(1.096)	(1.189)	(0.980)	(1.096)	(1.188)
No.(Sub-advisors in Fund)$_{i,t-1}$	0.344	0.238	0.142	0.346	0.239	0.143
	(0.383)	(0.294)	(0.254)	(0.382)	(0.294)	(0.254)
Constant	-17.794***	-19.342***	-18.504***	-17.782***	-19.335***	-18.497***
	(4.489)	(5.047)	(5.095)	(4.487)	(5.046)	(5.094)
Observations	30,000	30,000	30,000	30,000	30,000	30,000
Number of Funds	0.097	0.129	0.135	0.098	0.129	0.135
R^2	284	284	284	284	284	284
Fund FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES

Chapter 4

Does the Replication Method Affect ETF Tracking Efficiencies?

Xinrui Zheng*

September 28, 2021

Abstract

Using 2,290 European equity and fixed income ETFs from 2001 to 2020, this paper studies how the replication method affects the tracking efficiencies of ETFs, especially during market crises. There is no persistent evidence suggesting superior tracking performance of synthetic ETFs relative to physically-replicated ones. I identify 119 indices simultaneously tracked by both physical and synthetic ETFs, and conduct a difference-in-difference analysis around Lehman Brothers bankruptcy, sovereign debt crisis, and COVID-19 outbreak. Synthetic ETFs face steeper declines in tracking efficiencies after a sudden increase in counterparty risk, but they are shielded from liquidity shocks. There is a remarkable drop in the sensitivity of tracking performance to market distress measures after the global financial crisis.

Keywords: Exchange-traded Fund (ETF), Synthetic Replication, Tracking Efficiency, Counterparty Risk, Liquidity, VIX, Distress, Principle Component Analysis.

JEL classification: G10, G11, G12

*Judge Business School, University of Cambridge, Address: Trinity College, Cambridge, UK CB2 1TQ, email: xz359@cam.ac.uk. I thank Pedro Saffi for insights and comments. I gratefully acknowledge full financial support from Cambridge Judge Business School (CJBS). All remaining errors are my own.

4.1 Introduction

There are two fundamentally different ways that exchange-traded funds (ETFs) can replicate their underlying benchmark indices, namely physical replication and synthetic replication.[1] Physical replication involves holding all constituent securities or a representative sample of the benchmark index. Synthetic replication achieves the benchmark return by entering into a total return swap or other derivative contract with a counterparty, typically a large investment bank. The first synthetic ETF was introduced on the French market in 2001. Since then, synthetic structures have become more popular in Europe than in the US due to different regulations on fund's legal structures.[2] In this paper, I examine all equity and fixed income ETFs in Europe from 01 January 2001 to 31 December 2020 to see if the replication method affects the tracking error.

Figure 4.1 shows the percentage of ETFs with synthetic replication over time. The market share of synthetic ETFs reaches its peak in 2010, representing 50% of the total number of ETF offerings in Europe and 30% of aggregate assets under management. However, despite the exponential growth in ETF assets, the market share of synthetic ETFs has been shrinking in the past decade. By the end of 2020, synthetic ETFs represent only 23% in number and 17% in total net asset (TNA) of European ETFs. After the global financial crisis, synthetic ETFs has been widely criticized by regulators and financial advisors for their complexity, lack of transparency and counterparty risk. However, there is a resurgence in interest for synthetic ETFs, particularly those providing exposure to the US equity market due to their tax advantage over their physical peers (Zarate et al., 2021).[3] It is therefore important to known whether fund

[1] Physical replication can be divided into full replication and sampling, synthetic replication can be based on either total return swaps or other derivatives, such as futures contracts. A more detailed classification can be found at: http://www.argos-tsp.com/en/research/argos-finneo/terminological-research/summary-of-index-replication-methods-used-by-etf-providers.html.

[2] Most US-registered ETFs are governed by the Investment Company Act 1940 (ICA), which prohibits transactions between fund and its affiliate as well as other forms of self-dealing. As a result, organizing an ETF using synthetic structure becomes complicated. On the other hand, the majority of European-listed ETFs are regulated by UCITS, which allows the use of exchange-traded as well as OTC derivatives to achieve investment objectives, and therefore synthetic replication becomes more popular in Europe.

[3] BlackRock, once being a major critic of the synthetic structure, has launched a swap-based S&P 500 UCITS ETF in September 2020. Here is the Financial Times article: https://www.ft.com/content/6600bd7f-5433-47d3-a2df-04411e6de75b.

families offering synthetic ETFs learn from market failures and improve the risk management of synthetic structures since then.

This paper addresses the concern by answering the following three questions. First, do synthetic ETFs posses superior tracking ability compared to physically-replicated ones? Second, which replication method can better withstand market distress? Third, is there any improvement on risk management after the global financial crisis, especially in terms of the swap counterparty risk of synthetic ETFs? In this paper, I find no evidence of persistent superior tracking ability of synthetic ETFs across the sample period, especially after controlling for heterogeneity across the investment objectives. There are significant cross-sectional variations in tracking errors. Furthermore, after the global financial crisis, I observe a large reduction in tracking errors. Synthetic ETFs face steeper declines in tracking efficiencies after a sudden increase in counterparty risk. But during liquidity shocks, their tracking ability is less affected relative to the physical ones. I explain the relative tracking performance between physical and synthetic replication around market crisis by describing the trade-off between counterparty and liquidity risk that dominates the market. Finally, I find that the tracking performance of both physical and synthetic ETFs becomes significantly less sensitive to market distress after the global financial crisis. Specifically, synthetic equity ETFs demonstrate superior tracking ability in terms of both lower tracking errors and lower sensitivity to market turbulence in the post-crisis period.

There are a handful of studies comparing the tracking efficiencies between physical and synthetic ETFs (Elia, 2011; Johnson et al., 2013; Meinhardt et al., 2014; Naumenko and Chystiakova, 2015; Mateus and Rahmani, 2017). They look at different exchanges across different sample periods and find contradicting results. In this paper, I use a sample of all European equity and fixed income ETFs from 2001 to 2020 with daily observations. Unlike the previous work, I account for the significant time-series and cross-sectional variations in tracking errors, which largely reconcile the inconsistencies in the literature. There is a structural break in the level of tracking errors after the 2008 global financial crisis, especially among equity ETFs. Cross-sectionally, funds with high return volatility and poor past performance are associated with higher tracking errors. Although a synthetic structure is applied more commonly in less liquid or less developed markets, I find no persistent relation between the liquidity or efficiency of an objective market and the relative tracking efficiencies.

To address the second research question, I identify 119 physical and synthetic ETF pairs that have the same underlying benchmark, and conduct a difference-in-difference (DiD) analysis. Previous literature documents higher tracking errors for all ETFs during crisis periods characterized by large bid-ask spreads, small trading volumes, and high volatility of currency and exchange rates (Buetow and Henderson, 2012; Johnson et al., 2013). In this paper, I disentangle the impact of extreme market movements on ETFs with different replication methods. Results from the DiD analysis suggest that equity (fixed income) ETFs with synthetic structures experience larger declines in tracking efficiencies around Lehman Brothers' bankruptcy (sovereign debt crisis). However, synthetic ETFs fared better during the COVID-19 market shock in 2020, compared to their physical equivalents. Therefore, I relate the different tracking ability between physical and synthetic ETF pairs during extreme market turbulence to the trade-off between liquidity risk and counterparty risk.

Finally, I answer the third question by collecting several state variables that capture market distress from different perspectives, ranging from stock market volatility to credit risk and liquidity. VIX is the CBOE volatility index, which measures the expectation of stock market volatility over the coming 30 days. $VSTOXX$ is the Euro Stoxx 50 volatility index, commonly known as the "Euro VIX", which measures the implied volatility of near term EuroStoxx 50 options. $NOISE$ is a market-wide liquidity measure proposed by Hu et al. (2013). TED is calculated as the spread between 3-month LIBOR based on US dollars and 3-month treasury bill. I also extract a common factor using principal component analysis (PCA) to proxy for the overall state of the market. I introduce a three-way interaction using an indicator for synthetic replication, an indicator for the post-crisis period, and the daily change in state variables to explore how tracking performance responds differently to market movements across time. I find that the tracking performance of both physical and synthetic ETFs becomes less sensitive to market distress after the global financial crisis. In particular, the post-crisis tracking ability of synthetic ETFs is markedly less affected by market turbulence, compared to physically-replicated ones. This finding suggests significant improvements in risk management over the past decade, especially regarding swap counterparty risk among synthetic ETFs.

This paper contributes to the literature from three perspectives. First, this paper disentangles the effect of market distress on ETF tracking efficiencies. Difference in tracking performance deterioration around market crisis between ETFs with physical and synthetic replication is attributed to their different reactions to counterparty risk and liquidity risk. Second, the extensive 20-year sample period allows identification of several physical and synthetic ETF pairs tracking the same underlying benchmarks, and therefore a like-for-like comparison across different types of major market crisis. To my best knowledge, this is the first paper to make a direct comparison of ETFs with different replication methods on the same underlying index.[4] Third, I identify a structural break in both the level and sensitivity of market-wide tracking efficiencies after the global financial crisis, and provide empirical evidence on how market failure evokes tighter risk management.

The rest of this paper is organized as follows. Section 2 explains the major hypotheses. Section 3 describes the data and shows some summary statistics. Section 4 explains the research design and elaborates the empirical results from the analysis. And Section 5 concludes.

4.2 Hypothesis Development

A major objective of this paper is to compare the tracking efficiencies between physical replication and synthetic replication. Tracking error arises from many sources, including transaction and rebalancing costs, cash drag, dividend distribution and reinvestment, taxation, security lending, etc (Johnson et al., 2013). Some factors are more relevant to a particular type of replication method among others. For instance, a physical ETF can be forced to trade the underlying securities during reconstitution of the benchmark index, and is therefore exposed to trading frictions and liquidity risk. Also, there are periods when a proportion of the portfolio needs to be held in cash, mostly due to index rebalancing or dividend distribution. Physical ETFs are more prone to tracking inefficiencies arising from these periods of cash drag. Whereas a synthetic ETF in theory provides a more precise replication of the underlying index through

[4] Meinhardt et al. (2014) explains the scarcity in physical and synthetic ETFs mimicking the same index.

total return swap or other derivative contract. This leads to the first major hypothesis:

Hypothesis 1 *Without the need to physically hold the underlying securities and rebalance the portfolio, synthetic ETFs posses better tracking ability and therefore lower tracking errors.*

In real-world practice, synthetic structures are often applied to grant access to less liquid assets or less efficient markets, which are associated with higher replication costs (Ramaswamy, 2011). Besides, many physical ETFs engage in security lending activities, which are less prevalent among synthetic ETFs (Hurlin et al., 2019). Since the securities post as collateral by the swap counterparties are normally less liquid compared to the securities held for physical replication (Ramaswamy, 2011). Security lending generates additional income to offset the expenses associated with running the fund, while at the same time brings in additional counterparty risk. The complexity and variety in the sources of tracking errors suggest significant cross-sectional variations in tracking errors.

Next, I investigate the reasons behind different tracking performance between physical ETFs and synthetic ETFs in the face of market crisis. On the one hand, liquidity dries up during market distress, buying and selling of the underlying securities becomes increasingly costly. Synthetic ETFs do not need to trade the underlying assets, and are therefore better shielded from liquidity shocks. However, the default probability of the swap counterparties for synthetic ETFs also increases sharply during extreme market downturns. The additional counterparty risk faced by synthetic ETFs could drive up the tracking errors. Though physical ETFs engaging in security lendings are also exposed to counterparty risk, the security lending programs usually cap the amount that could be loaned out and the transactions are often overcollateralised by 10-20% (Ramaswamy, 2011).[5] I posit that the relative tracking performance during market dislocations is determined by a trade-off between liquidity risk and counterparty risk.

In the equity universe, the bankruptcy of Lehman Brothers in September 2008 is associated with substantial counterparty risk (Fender and Gyntelberg, 2008; Korajczyk and Sadka, 2008).

[5] Although there is no bespoke regulations stating the maximum percentage of portfolio holdings that can be lent out, UCITS ETFs must disclose both the maximum and expected percentages of securities lending usage in its offering documentation. https://kraneshares.eu/breaking-down-securities-lending-benefits-to-etf-investors/.

In the fixed income universe, the market experienced the greatest level of counterparty risk during the 2011 sovereign debt crisis (BIS, 2011). While the 2020 market crash caused by the outbreak of COVID-19 pandemic is mostly dominated by liquidity risk across both equity and fixed income markets, without raising major concerns over counterparty defaults (ECB, 2020). This leads to the next hypothesis:

Hypothesis 2 *Equity (fixed income) ETFs with synthetic structures would experience steeper decline in tracking efficiencies around Lehman Brothers bankruptcy (sovereign debt crisis). Meanwhile, the tracking efficiencies of physical ETFs would be more strongly affected by the liquidity shocks around the COVID-19 outbreak in both equity and fixed income markets.*

Losses experienced during the global crisis lead to better awareness and tighter management of counterparty risk thereafter (Grill et al., 2018). Zarate et al. (2021) point out that the use of multiple swap counterparties has become more common among synthetic ETFs to mitigate swap counterparty risk in the recent decade. Also, almost all swap-based ETFs nowadays apply the unfunded structure.[6] Under unfunded swap structure, the ETF sponsor is the direct beneficial owner of the collateral assets, which avoids the potential delay in realising the net asset value (NAV) of the collateral in the case of counterparty default (Ramaswamy, 2011). Transparency over the disclosure of collateral baskets has improved evidently compared to a decade ago.[7] There are also regulatory efforts spent on imposing additional requirements over the quality and liquidity of collaterals from third parties (ESMA, 2012). With more stringent standards on risk management, the tracking efficiencies are expected to be less dependent on market movements and ETFs could better withstand market crisis. This leads to the third hypothesis:

Hypothesis 3 *Tracking errors become less sensitive to market distress post the global financial crisis, especially for synthetic equity ETFs.*

[6] According to Zarate et al. (2021), UBS is the only one that applies funded swap structures among all major ETF providers.

[7] Full details regarding the constituents of the collateral baskets are publicized online and typically updated on a daily basis by the ETF providers (Zarate et al., 2021).

4.3 Data and Sample

4.3.1 Sample Selection

In this study, I started with a sample of 3,550 equity and fixed income ETFs provided by 79 different fund families, whose primary share class is listed on an European stock exchange from 1st January 2001 to 31th December 2020.[8] Fund characteristics including TNAs, daily returns, prices, spreads and volumes, as well as the annual report expense ratios are obtained from Morningstar Direct Database. Both surviving and delisted funds are included in the sample to avoid survivorship bias. Replication method is obtained from Morningstar, complemented by information disclosed about the fund holdings. Funds without disclosure about the replication strategy or where the replication method is not applicable, such as active funds and fund-of-funds, are excluded from the sample.[9] Information on the primary prospectus benchmark is manually matched with Bloomberg and the official websites of ETF providers. Only the fund observations with matched benchmark indices and non-missing time series of daily returns are kept. The tracking efficiencies of leveraged and inverse ETFs are complicated by compounding effect on top of the replication strategy (Shum and Kang, 2012). I therefore exclude them from the sample. The final sample consists of 2,290 funds in total with 1,508 equity funds and 782 fixed income funds. On average, 48% of equity ETFs and 7% of fixed income ETFs use synthetic replication.

All funds in the final sample are classified into 13 equity categories and 7 fixed income categories, as listed in Table 4.1. Based on the investment objective indicated by the primary prospectus benchmark, all equity funds are divided into two streams, namely specialized and broad-based (Ben-David et al., 2021). A specialized fund has a specific focus on a particular sector or theme, for instance industrial, infrastructure, energy, Technology, etc. A broad-based

[8]Exchange-traded notes (ETNs) and exchange-traded commodities (ETCs) are excluded from the original sample. Morningstar uses "ETF" as an umbrella term to refer to a range of different ETPs, including ETFs, ETNs and ETCs. Unlike ETFs, ETNs and ETCs are more accurately described as debt securities that are less relevant to our study.

[9]Full replication and sampling strategy are both classified as physical replication. Full replication requires physically holding all underlying securities included in the index, while in physical sampling, an optimized portfolio of securities are selected to represent the index based on correlations, exposure and risk.

fund tracks a broad-market index. I further categorize them according to the geographic focus and investment style. Fixed income funds are categorized straightforwardly according to the investment area.

In order to explore the relation between tracking efficiencies and market distress, I collect four state variables to capture different aspects of the overall market condition. The first state variable is the CBOE volatility index (VIX) obtained from FRED, which measures the expectation of stock market volatility over the coming 30 days.[10] Since the investment mandate of 25% equity funds and 46% fixed income funds in our sample has a European (incl. UK) focus, I include the Euro Stoxx 50 volatility index (VSTOXX) obtained from Qontigo as a second state variable.[11] VSTOXX is commonly known as the "Euro VIX", which measures the implied volatility of near term EuroStoxx 50 options. I also include a market-wide liquidity measure NOISE proposed by Hu et al. (2013).[12] The final state variable is a measure of credit risk, namely TED spread, obtained from FRED.[13] TED spread is calculated as the spread between 3-month LIBOR based on US dollars and 3-month Treasury Bill. On top of the above four continuous state variables, I also include an indicator variable *Recession* for US market recession periods derived from business cycle turning points determined by the National Bureau of Economic Research (NBER).[14] *Recession* takes a value of one during three sample periods, namely March 2001 to November 2001, December 2007 to June 2009 and Febuary 2020 to April 2020.

[10]Chicago Board Options Exchange, CBOE Volatility Index: VIX [VIXCLS], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/VIXCLS,September23,2021.

[11]Historical daily index level of VSTOXX is available at www.stoxx.com/document/Indices/Current/HistoricalData/h_vstoxx.txt.

[12]Historical daily measure of NOISE can be downloaded from http://en.saif.sjtu.edu.cn/junpan/.

[13]Federal Reserve Bank of St. Louis, TED Spread [TEDRATE], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/TEDRATE,September22,2021.

[14]The NBER recession data are available at http://www.nber.org/cycles/cyclesmain.html.

4.3.2 Variable Definitions

Tracking errors are measured in four different ways throughout the sample period. At daily frequency, I follow Frino and Gallagher (2001) and compute the absolute tracking difference:

$$TE\_abs_{i,t} = |R_{i,t} - R_{BMK_i,t}|. \tag{4.1}$$

At monthly frequency, three additional measures of tracking errors are calculated from daily returns. The most commonly applied measure among practitioners is the standard deviation of fund excess returns over the benchmark (Rompotis, 2005):

$$TE\_sd_{i,k} = \sqrt{\frac{\sum_{t=1}^{N}\left((R_{i,t} - R_{BMK_i,t}) - \overline{(R_{i,t} - R_{BMK_i,t})}\right)^2}{N-1}}. \tag{4.2}$$

Shin and Soydemir (2010) substitute the excess returns by the absolute return differences:

$$TE\_sd\_abs_{i,k} = \sqrt{\frac{\sum_{t=1}^{N}\left(|R_{i,t} - R_{BMK_i,t}| - \overline{|R_{i,t} - R_{BMK_i,t}|}\right)^2}{N-1}}. \tag{4.3}$$

The last measure is to compute the standard error of residuals within month k (Frino and Gallagher, 2001; Shin and Soydemir, 2010):

$$TE\_se\_res_{i,k} = \sqrt{\frac{\sum_{t=1}^{N}\varepsilon_t^2}{N-2}}, \tag{4.4}$$

where the residuals ε_t are derived from the following regression:

$$R_{i,t} = \alpha_i + \beta_{i,t}R_{BMK_i,t} + \varepsilon_{i,t}. \tag{4.5}$$

Pope and Yadav (1994) show that serial correlation in returns could bias the estimates of tracking errors. Durbin-Watson test suggests significant first-order auto-correlation in the above regression residuals. Serial correlation leaves the coefficient estimates unbiased but distorts the standard errors and thus the efficiency of statistical inferences. Newy-West standard errors are applied to relieve the distortion from serial correlation in daily returns.

To measure the aggregate risk level of a fund, return volatility is calculated as the standard deviation of daily excess returns over the past 21 trading days. Benchmark Alpha is calculated as the difference between the daily realized return and the predicted return with beta estimated from the single index model using a rolling window of prior 252 trading days:

$$R_{i,t} - r_f = \alpha_i + \beta_{i,t}(R_{BMK_i,t} - r_f) + \varepsilon_{i,t}, \tag{4.6}$$

where $(R_{BMK_i,t} - r_f)$ denotes the excess return of the primary benchmark over the risk-free rate for fund i on day t. Daily dollar volume is computed by multiplying the daily volume and the daily trading price.

4.3.3 Descriptive Statistics

Table 4.2 reports the descriptive statistics for all four tracking error measures and major fund characteristics, including return volatility, benchmark alpha, TNA, age, expense ratio, bid-ask spread and dollar volume. Summary statistics on the daily change in levels of VIX, VSTOXX, NOISE and TED spread are also reported. In panel A, summary statistics are reported for all funds in the sample across the 20-year sample period 01 January 2001 to 31 December 2020. Panel B compares between equity and fixed income ETFs. Panel C compares between ETFs with physical replication and synthetic replication. The "MeanDiff" column in Panel B and C reports the difference in the sample means, together with the statistical significance from a two-sided t-test. The mean tracking error derived from daily returns in the full sample ranges from 0.12% to 0.20% across four different measures, the median ranges from 0.01% to 0.23%. The magnitude is consistent with the range from 4 to 7bps for annualised tracking error found by Johnson et al. (2013). On average, an ETF underperforms its benchmark index by 1.64bps per month. A median fund in the sample has $141 million asset under management, charging a total fee of 0.33% per annum. The median daily bid-ask spread is 0.1% and the median daily dollar volume is $0.03 million.

Panel B indicates that equity ETFs on average have smaller size compared to fixed income ETFs, and experience significantly higher return volatility and daily dollar volume. The average expense ratio charged by equity ETFs are 0.20% higher than fixed income ETFs, which is economically substantial. Daily absolute tracking differences of equity ETFs are on average 5bps lower than fixed income ETFs. While tracking errors measured at monthly frequency suggest marginally higher tracking efficiencies of fixed income ETFs over equity. Panel C indicates that synthetic ETFs on average have smaller size and age compared to physical ETFs. Synthetic ETFs also experience significantly larger return volatility, bid-ask spread and dollar volume. On average, they charge 6bps higher annual expense ratios and underperform physical ETFs by 13bps daily. Tracking errors across all four measures at both daily and monthly frequency consistently point to lower tracking efficiencies of synthetic ETFs. However, a rough comparison over the entire sample could mask significant variations in tracking errors across investment objectives and through time, which is unveiled in the next section.

4.4 Results

4.4.1 Comparing Tracking Efficiencies between Physical and Synthetic Replications

In this section, I examine Hypothesis 1 and compare the tracking efficiencies between physical ETFs and synthetic ETFs within different category groups and across different sample periods. Figure 4.2 plots the aggregate tracking errors for physical ETFs versus synthetic ETFs separately within the equity universe and fixed income universe. The times series expands the whole sample period from 01 January 2001 to 31 December 2020, with shaded areas indicate the periods of major market crises.[15] In the equity world, the global financial crisis brings the highest level of market turbulence, reaching a climax at the bankruptcy of Lehman Brothers on

[15] The first synthetic ETF was an equity ETF introduced into the French market in 2001. While the first synthetic fixed income ETF in our sample only appears in 2007.

16 September 2008.[16] The NBER declared June 2009 as the end date of the U.S. recession.[17] The shaded period expands from September 2008 to June 2019 in accordance with the NBER market turning points. In the fixed income universe, the sovereign debt crisis originated from eurozone government deficits from April 2010 to October 2012 has the most severe impact on market stability. I therefore identify the post-crisis period to be the sample period after June 2009 for equity funds and after October 2012 for fixed income funds. The time series suggests that the average tracking error is noticeably smaller in the post-crisis period for both physical and synthetic ETFs, especially in the equity universe. This confirms the premise that there exist significant time-series variations in ETF tracking efficiencies.

Recognizing the structural break in the level of tracking errors post crisis, I conduct a repeated measure ANOVA analysis to compare the sample means between physical and synthetic ETFs, sorted into four mutually exclusive groups by asset classes and sample periods. Figure 4.3 shows the ANOVA plots of group means with the 95% confidence intervals from Tukey's Post Hoc statistics. In the pre-crisis period, the mean tracking error of fixed income ETFs is indifferent between physical and synthetic replications across all four measures. While the pre-crisis tracking error of synthetic equity ETFs are on average 0.6% - 1.0% higher than physical equity ETFs across different measures, which is significant both statistically and economically. Post crisis, the relation is reversed. The average tracking error of both equity and fixed income ETFs with synthetic replication is marginally smaller than ETFs with physical replication. Together, the above patterns suggest that the significant tracking difference shown in Panel C of Table 4.2 is dominated by equity funds pre-crisis. This also reconciles the discrepancies from different findings. For instance, Mateus and Rahmani (2017) find no evidence on superior tracking performance of synthetic equity ETFs traded on the London Stock Exchange (LSE), as the sample period 2008-2013 is distorted by the extreme deterioration in tracking efficiencies of synthetic ETFs before June 2009. While Johnson et al. (2013) finds lower tracking error on ETFs using synthetic replications in the case of seven out of eight benchmarks studied during the period 2010-2012, which is consistent with the pattern for post-crisis equity group in the

[16] https://en.wikipedia.org/wiki/Financial_crisis_of_2007_2008.

[17] The FOMC statement is available at: https://www.federalreserve.gov/newsevents/pressreleases/monetary20090624a.htm.

above ANOVA analysis.

Next, I investigate the cross-sectional variations in tracking efficiencies. Table 4.1 reports the category distribution of synthetic ETFs and compares the tracking efficiencies between two replication methods within each individual investment objective. This table allows us to explore if synthetic structure is more commonly applied to less liquid assets or to grant access to remote markets, and therefore associated with higher costs and higher tracking errors. In the equity universe, synthetic structure is more frequently applied in sector equity (60%), Africa & Latin America equity (59%), emerging markets equity (50%), followed by US equity mid/small cap (48%) and UK equity mid/small cap (45%). These categories indeed corresponds to less liquid assets and remote markets as opposed to US/Europe equity large cap among others. However, there is no persistent relation between the liquidity and efficiency of an objective market and the relative tracking efficiencies. In 3 out of 5 categories listed above, synthetic ETFs show inferior tracking performance compared to physical ETFs. While in 6 out of 11 equity categories, synthetic ETFs demonstrate superior tracking abilities. In the fixed income universe, synthetic structure is most commonly applied in European (incl.UK) products, with a frequency at around 20%.[18] Across the whole sample period, synthetic funds within these categories exhibit slightly inferior tracking abilities.

Overall, there is no persistent evidence on either superior or inferior tracking ability of synthetic ETFs over physical ETFs across the whole sample, i.e. Hypothesis 1 is rejected. There exists a structural break in the aggregate level of tracking errors post crisis. Though synthetic structure is more common in less liquid or less efficient markets, this does not translate into higher tracking errors directly.

4.4.2 Factors Affecting ETF Tracking Efficiencies

In the previous section, I compare the tracking efficiencies between physical and synthetic ETFs in different sample periods and across different investment objectives. In this section, I study the time-series and cross-sectional variations in tracking errors simultaneously under a

[18]There are 2 out of 5 synthetic funds in the fixed income miscellaneous group, which we do not discuss in here.

pooled panel regression framework. The regression model is specified as:

$$\begin{aligned} TE\_abs_{i,t} &= \beta_0 + \beta_1 Synthetic * I + \beta_2 Synthetic + \beta_3 I \\ &+ \beta_4 Return\ Volatility_{i,t-1} + \beta_5 Benchmark\ Alpha_{i,t-1} + \beta_6 Log(TNA)_{i,t-1} \\ &+ \beta_7 Age_{i,t-1} + \beta_8 Expense\ Ratio_{i,t-1} + \beta_9 Bid\text{-}Ask\ Spread_{i,t-1} \\ &+ \beta_{10} Dollar\ Volume_{i,t-1} + \gamma_t + \varepsilon_{i,t}, \end{aligned} \tag{4.7}$$

where the dependent variable $TE\_abs_{i,t}$ is the daily absolute tracking difference. *Synthetic* is an indicator variable taking a value of one if fund i is under synthetic replication, and zero otherwise. I represents an indicator for equity funds in the full sample, and an indicator for the post-crisis period in the equity and fixed income subsamples. The *Post Crisis* indicator marks the period June 2009 to December 2020 in the equity subsample and October 2012 to December 2020 in the fixed income subsample, as explained in the previous section. Return volatility is measured from a 21-trading day rolling window, all other explanatory variables are lagged one trading day. In the full sample, time fixed effect is included as γ_t. In the equity and fixed income subsamples, category fixed effect is applied instead to account for heteroskedasticity across different investment objectives.[19] Pope and Yadav (1994) as well as Meinhardt et al. (2014) show that the existence of serial correlation in daily returns has significant impact on measures of tracking errors. A portmanteau test for fixed effects models proposed by Inoue and Solon (2006) is applied to test for serial correlation in residuals, and the null of independent error terms is rejected at 1% level. Therefore, Prais-Winsten transformation is applied to the residuals (Prais and Winsten, 1954). All variables are standardized to have a mean of zero and standard deviation of one, to allow for more direct interpretation and comparison of coefficients.

The regression results are reported in Table 4.3. In the full sample, neither being an equity fund or synthetic fund is associated with any significant difference in tracking errors. In the equity subsample, synthetic funds are associated with 10.30 $\left(= \dfrac{2.352 * 0.552}{0.126}\right)$ times higher

[19] Fund and category fixed effect is not included in the full sample as the two indicator variables of our interest, namely *Synthetic* and *Equity* are invariant within funds or categories. Similarly, time fixed effect is not included in the equity and fixed income subsamples due to the inclusion of the *Post Crisis* indicator.

daily absolute tracking differences relative to the mean before June 2009, but the tracking error falls slightly below physical funds post crisis.[20] The post-crisis period is associated with $127\% = \dfrac{0.289 * 0.552}{0.126}$ $\left(1,216\% = \dfrac{(0.289 + 2.488) * 0.552}{0.126}\right)$ lower daily absolute tracking differences relative to the mean for physical (synthetic) ETFs. These results confirm significant variations in tracking efficiencies across different time periods.

The panel regression results also confirm significant cross-sectional variations in tracking efficiencies. Across all three models, the riskiness of fund returns exhibits a strong and positive relation with the tracking error. In the full sample, a one standard deviation increase in fund return volatility over the past month is associated with a 0.034 standard deviation increase in the daily absolute tracking difference, which is equivalent to a 14.9% $\left(= \dfrac{0.034 * 0.514}{0.117}\right)$ increase relative to the mean tracking difference. While past performance is negatively related related to the tracking error across all three models. A one standard deviation increase in the lagged *Benchmark Alpha* is associated with a 0.021 standard deviation decrease in daily absolute tracking difference, which is equivalent to a 9.2% $\left(= \dfrac{0.021 * 0.514}{0.117}\right)$ decrease relative to the mean tracking difference. This is consistent with the risk-shifting behaviour of fund managers following poor past performance (Chevalier and Ellison, 1997). There is also a positive relation between fund size and the tracking difference, which is more prominent in the fixed income subsample. This could be partly due to the difficulty for a large fund to effectively track the underlying without significantly move the securities prices (Pástor et al., 2015; Magkotsios, 2018). In the equity subsample, younger funds are associated with higher tracking errors, the relation is not found in the fixed income subsample. In the fixed income model, a one standard deviation increase in the expense ratio is associated with a 0.05 standard deviation increase in the daily absolute tracking difference, which is equivalent to a 18.2% $\left(= \dfrac{0.050 * 0.262}{0.072}\right)$ increase relative to the mean tracking difference.[21] This confirms the negative impact of total cost on tracking efficiencies found by Frino and Gallagher (2001), Chu (2011) and Johnson et al. (2013). However, the relation does not hold in the equity subsample. Moreover, the liq-

[20]The percentage decrease is calculated using statistics for the equity subsample, where the daily TE_abs has a mean of 0.126% and a standard deviation of 0.552%.

[21]The percentage increase is calculated using statistics for the fixed income subsample, where the daily TE_abs has a mean of 0.072% and a standard deviation of 0.262%.

uidity measures *Bid-Ask Spread* and *Dollar Volume* show no significant impact on tracking errors after controlling for category fixed effects. This contradicts the intuition that synthetic structure being more widely applied in less liquid markets is associated with higher tracking errors.

4.4.3 Relative Tracking Efficiencies of ETF Pairs around Crisis

In this section, I examine Hypothesis 2 and investigate the differential impact of major market crises on tracking efficiencies between physical and synthetic ETFs. Recognizing the existence of significant cross-sectional variations in tracking errors, I construct a subsample of physical and synthetic ETF pairs tracking the same underlying benchmark to enable a like-for-like comparison. The whole sample consists of 2,290 funds tracking 957 distinct benchmarks. There are 528 benchmark indices tracked by a unique fund throughout the sample period, the other 429 benchmark indices have more than one funds tracking.[22] I identified 119 benchmarks which are followed by both physical and synthetic ETFs simultaneously. In the cases where there are more than one physical/synthetic ETFs tracking the benchmark at the same time, the daily absolute tracking difference of each fund with the same replication structure is equally weighted.

Figure 4.4 plots the average daily absolute tracking differences for physical and synthetic ETF pairs following the same underlying benchmark in the [-21,+126] trading days event window around major market crises. Among equity ETF pairs, the tracking error of synthetic funds rocketed significantly more compared to physical funds following Lehman Brothers bankruptcy, while they reacted less dramatically following the outbreak of COVID-19. Similar pattern is also present among fixed income ETF pairs. The tracking error of synthetic ETFs went up significantly more than their physical equivalents during the sovereign debt crisis, especially within the first two months, which is in stark contrast to their milder reaction following COVID-19 outbreak.

[22] S&P 500 and Euro Stoxx 50 are the two most popular benchmarks followed by 41 and 33 funds in the sample, respectively. MSCI USA and MSCI World are both followed by 29 funds during the sample period 2001-2020, ranking the third.

I conduct a difference-in-difference (DiD) analysis on the tracking performance between ETF pairs in a event window of [-21,+126] trading days around major market crisis. The baseline regression model follows Equation (4.7), with the indicator variable I being replaced by a vector of indicators representing different post-event time windows. As before, category fixed effect is included. Prais-Winsten transformation is applied to the residuals (Prais and Winsten, 1954). All variables are standardized to have a mean of zero and standard deviation of one. The DiD analysis is applied to the equity and fixed income subsample separately. To study the trade-off between liquidity risk and counterparty risk in determining relative tracking efficiencies, I identify the bankruptcy of Lehman Brothers on 15 September 2008 as a representative of sudden increase in counterparty risk affecting the equity market. In a similar vein, the Sovereign Debt Crisis on 27 April 2010 is considered an unexpected surge in counterparty risk hitting the fixed income market. While the outbreak of COVID-19 on 20 Febuary 2020 mainly leads to rise in liquidity risk in both equity and fixed income markets, without raising major concerns about counterparty risk.

Table 4.4 reports the results from the above DiD analysis. In Panel A, the marginal decline in tracking efficiencies of equity ETF pairs is contrasted in the post-event windows between Lehman Brothers bankruptcy (Model (1)& (2)) and COVID-19 outbreak (Model (3)& (4)). The most interesting result is that the tracking performance of synthetic ETFs post Lehman Brothers bankruptcy drops significantly more than the physical equivalents within the same benchmark pair. While the decline in tracking efficiencies of synthetic ETFs post COVID-19 outbreak is significantly smaller than physical ETFs. According to Model (1), daily absolute tracking difference of synthetic ETFs in the first week, i.e. [+1,+5] trading days, post Lehman Brothers bankruptcy grows 122% $\left(=\dfrac{0.581*2.579}{1.228}\right)$ (relative to the mean) more than their physical equivalents. This DiD effect is on top of the fact that tracking error of synthetic ETFs pre-event is 63.8% $\left(=\dfrac{0.304*2.579}{1.228}\right)$ (relative to the mean) higher than physical ETFs. The incremental difference in tracking errors reaches a peak of 165% $\left(=\dfrac{0.785*2.579}{1.228}\right)$ (relative to the mean) in the first month, i.e.[+6,+21] trading days, post event. And the pattern persists at least up until 6 months post event. In Model (2), all continuous explanatory variables in Table 4.3 are included as controls. The incremental difference remains positive and is significant within the first month post event. In both models, there is also an increase in daily absolute tracking difference of physical equity ETFs in the 6-month window post crisis, though sometimes insignificant. The statistical pattern is reversed around the COVID-19 outbreak. According to Model (3), coefficients on the three event-window indicators are all positive and significant at 1% level, suggesting the tracking error of physical ETFs grows substantially larger post event. However, coefficients on the interaction terms are negative and significant in the [+6,+126] event window, indicating that the decline in tracking efficiencies due to COVID-19 outbreak is less severe for synthetic ETFs.

DiD analysis on fixed income ETF pairs exhibits similar patterns and the results are reported in Panel B of Table 4.4. Control variables are not included due to the lack of complete observations. There is no significant difference in tracking efficiencies between the ETF pairs pre-event. Around the sovereign debt crisis, the daily absolute tracking difference of synthetic ETFs grows 2.30 $\left(= \frac{0.658 * 0.028}{0.008}\right)$ times (relative to the mean) more than their physical equivalents during the [+6,+21] event window. While following the COVID-19 outbreak, the increase in tracking errors of synthetic ETFs is 63.6% $\left(= \frac{0.398 * 0.297}{0.186}\right)$ (relative to the mean) lower than the physical equivalents during the same [+6,+21] event window.

Together, results from the DiD analysis on ETF pairs in both equity and fixed income subsample confirm Hypothesis 2. The relative tracking performance between physical ETFs and synthetic ETFs during extreme market turbulence is determined by a trade-off between liquidity risk and counterparty risk. When the perception of counterparty risk hits the market unexpectedly, ETFs with synthetic replication would face steeper decline in their tracking efficiencies. While during liquidity shocks, the tracking performance of synthetic ETFs is better protected compared to the physical ETFs.

4.4.4 Sensitivity of Tracking Efficiencies to Market Distress

In this section, I examine Hypothesis 3 and explore how tracking performance sensitivity to market movements varies through time. In order to do so, I introduce a three-way interaction of the *Synthetic* indicator, the *Post Crisis* indicator and the lagged daily change in *State Variable* to investigate how synthetic ETFs in the post-crisis period respond to changes in the state variable differently. The three corresponding pair-wise interactions are also included and the regression model is specified as:

$$\begin{aligned}
TE_abs_{i,t} \quad = \quad & \beta_0 + \beta_1 Synthetic * Post\ Crisis * \Delta State\ Variable_{i,t-1} \\
& + \beta_2 Synthetic * \Delta State\ Variable_{i,t-1} + \beta_3 Post\ Crisis * \Delta State\ Variable_{i,t-1} \\
& + \beta_4 Synthetic * Post\ Crisis + \beta_5 Synthetic + \beta_6 Post\ Crisis \\
& + \beta_7 \Delta State\ Variable_{i,t-1} + \beta_8 Return\ Volatility_{i,t-1} \\
& + \beta_9 Benchmark\ Alpha_{i,t-1} + \beta_{10} Log(TNA)_{i,t-1} \\
& + \beta_{11} Age_{i,t-1} + \beta_{12} Expense\ Ratio_{i,t-1} + \beta_{13} Bid\text{-}Ask\ Spread_{i,t-1} \\
& + \beta_{14} Dollar\ Volume_{i,t-1} + \theta_i + \varepsilon_{i,t},
\end{aligned} \tag{4.8}$$

where the lagged daily change in VIX index, VSTOXX index, NOISE measure and TED spread are used in turn as $\Delta State\ Variable_{i,t-1}$, as well as a common factor $PC1$ derived from all four state variables using Principal Component Analysis (PCA). In the equity subsample, I also replace $\Delta State\ Variable_{i,t-1}$ with an indicator variable *Recession* marking the NBER defined US recession periods. As before, category fixed effect is included. Prais-Winsten transformation is applied to the residuals (Prais and Winsten, 1954). All variables are standardized to have a mean of zero and standard deviation of one.

Each of the four state variables measures the degree of market distress from a different perspective, ranging from stock market volatility to credit risk and liquidity. I follow the practice of Hasbrouck and Seppi (2001), Johnson (2008) and Richardson et al. (2017) and use PCA to extract a common factor to proxy for the state of the market.[23] Table 4.5 reports the results from the analysis. Panel A lists the eigenvalues and proportion of variance explained by the four principle components using 4,860 days when all state variables are jointly available. The first principle component ($PC1$) explains almost 70% of the total variance with an eigenvalue

[23] Hasbrouck and Seppi (2001) and Johnson (2008) apply PCA to extract common factors in stock returns, order flows and liquidity proxies. Richardson et al. (2017) use PCA to construct a common factor for funding liquidity measures.

of 2.78, being the only factor with eigenvalue above one. According to the Kaiser rule, only $PC1$ is selected. Panel B displays the eigenvectors, i.e. factor loadings, on the principle components. Column 1 shows that the four state variables contribute evenly to $PC1$, with loadings ranging from 0.409 to 0.558. In Panel C, I examine the correlation between $PC1$ and the four state variables. $PC1$ appears to be positively correlated (with correlation coefficients being above 0.60) with all four variables. The pairwise correlation coefficient between $PC1$ and VIX is 0.931.

In Table 4.6, the above regression (4.8) is run in the equity subsample. Control variables are included in all models, though not reported for brevity. In Model (1), synthetic ETFs during market recession after June 2009 face $13.3 = \frac{(1.718 + 0.352 + 0.840 + 0.130) * 0.552}{0.126}$ times (relative to the mean) less drop in tracking efficiencies compared to themselves during market recession before June 2009. For physical ETFs, the drop in tracking efficiencies is $2.1 = \frac{(0.352 + 0.130) * 0.552}{0.126}$ times (relative to the mean) less during market recession after June 2009. Especially in the post-crisis period, synthetic ETFs on average experience $94.2\% = \frac{(1.718 - 1.634 + 0.840 - 0.709) * 0.552}{0.126}$ (relative to the mean) less drop in tracking efficiencies than physical ETFs during market recession. The statistical pattern persists when the *Recession* indicator is substituted with changes in continuous state variables. Model (4) displays the results using the daily change in $PC1$ constructed above. The increase in daily absolute tracking differences for synthetic ETFs associated with a one standard deviation increase in $\Delta PC1$ in the post-crisis period is $12.3 = \frac{(0.029 + 0.013 + 2.469 + 0.290) * 0.552}{0.126}$ times (relative to the mean) less than the pre-crisis period. For physical ETFs, the increase in tracking differences is $1.3 = \frac{(0.013 + 0.290) * 0.552}{0.126}$ times (relative to the mean) less post crisis. Synthetic ETFs in the post crisis period experience $53.4\% = \frac{(0.029 - 0.039 + 2.469 - 2.337) * 0.552}{0.126}$ (relative to the mean) less increase in tracking differences than physical ETFs, when $\Delta PC1$ increased by one standard deviation. In Table 4.7, the same regression (4.8) is applied to the fixed income subsample. The statistical inference remains in the same direction across all model specifications, though the significance of results is diminished.

Overall, the results confirm Hypothesis 3. The sensitivity of tracking performance to market distress for both physical and synthetic ETFs drops significantly post the global financial crisis. In particular, the post-crisis tracking ability of synthetic ETFs is markedly less affected by market turbulence, compared to physical ETFs. This remarkable change in tracking performance sensitivity could partly due to closer scrutiny and better awareness of counterparty risk that market participants learnt from the disastrous global financial crisis.

4.5 Conclusion

This paper studies how the replication method affects the tracking efficiencies of ETFs, especially during market crises. There is no persistent evidence on either superior or inferior tracking ability of synthetic ETFs over physical ETFs across the whole sample. I first examine the relative tracking efficiencies across different objective markets and different sample periods. Synthetic structure is more common in less liquid or less efficient markets, however this does not translate into higher tracking errors directly. There exists a structural break in the aggregate level of tracking errors after the global financial crisis. Besides, there are significant cross-sectional variations in tracking differences. Higher return volatility, lower past performance, larger fund size, younger age, higher expense ratios are all associated with higher average tracking errors. Liquidity and volume do not show any significant impact on tracking efficiencies.

Next, I investigate how tracking efficiencies of physical and synthetic ETFs react differently around major market crisis. To enable a like-for-like comparison, I identify physical and synthetic ETF pairs tracking the same underlying benchmark and perform DiD analysis under an event-study framework. The relative tracking performance between physical ETFs and synthetic ETFs around crisis is determined by a trade-off between liquidity risk and counterparty risk. When the perception of counterparty risk hits the market unexpectedly, ETFs with synthetic replication face steeper decline in their tracking efficiencies. While during pure liquidity shocks, the tracking performance of synthetic ETFs is better protected compared to physical ETFs.

Finally, I explore how tracking performance responds differently to market movements across time. I find that sensitivity of tracking performance to market distress for both physical and synthetic ETFs drops significantly post the global financial crisis. In particular, the post-crisis tracking ability of synthetic ETFs is markedly less affected by market turbulence, compared to physical ETFs. This finding demonstrates significant improvements in risk management over the past decade post crisis. Synthetic ETFs has done particularly well in controlling counterparty risk.

Figure 4.1. Percentage of Synthetic ETFs through Years

This figure shows the percentage of ETFs with synthetic replication in each year from 2001 to 2020. The upper panel plots the relative number of fund offerings, and the lower panel plots the aggregate asset under management for funds with physical replication and synthetic replication respectively.

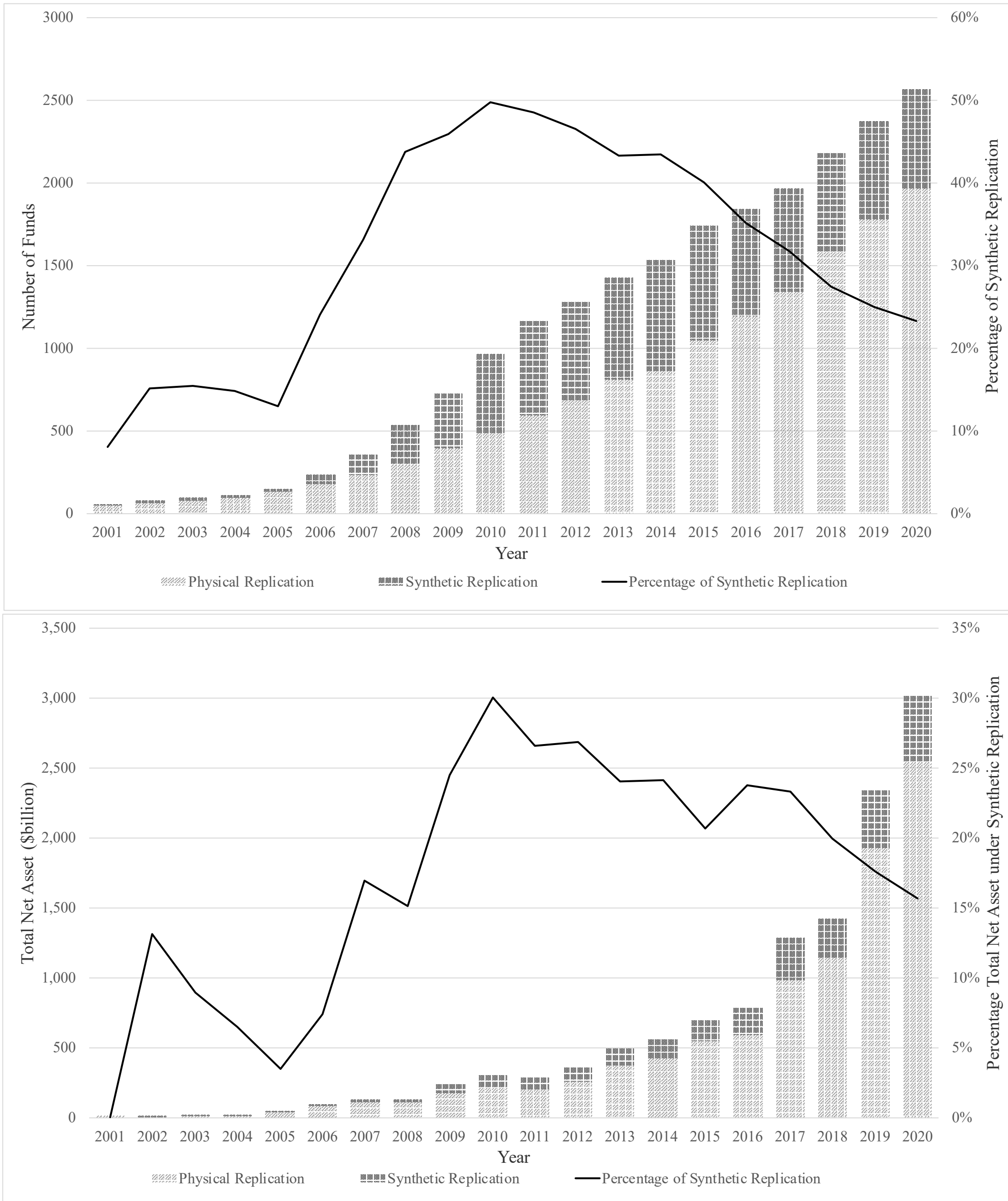


Figure 4.2. Time Series of Aggregate Tracking Errors

This figure plots the time series of aggregate tracking errors for physical and synthetic ETFs from January 2001 to December 2020. The tracking error is measured monthly as the standard deviation of daily excess returns over the benchmark. The upper panel displays the time series for equity funds, with the shaded area denoting the NBER defined market recession periods around Lehman Brothers bankruptcy and COVID-19 outbreak. The lower panel shows the time series for fixed income funds, with the shaded area denoting the NBER defined market recession periods around sovereign debt crisis and COVID-19 outbreak.

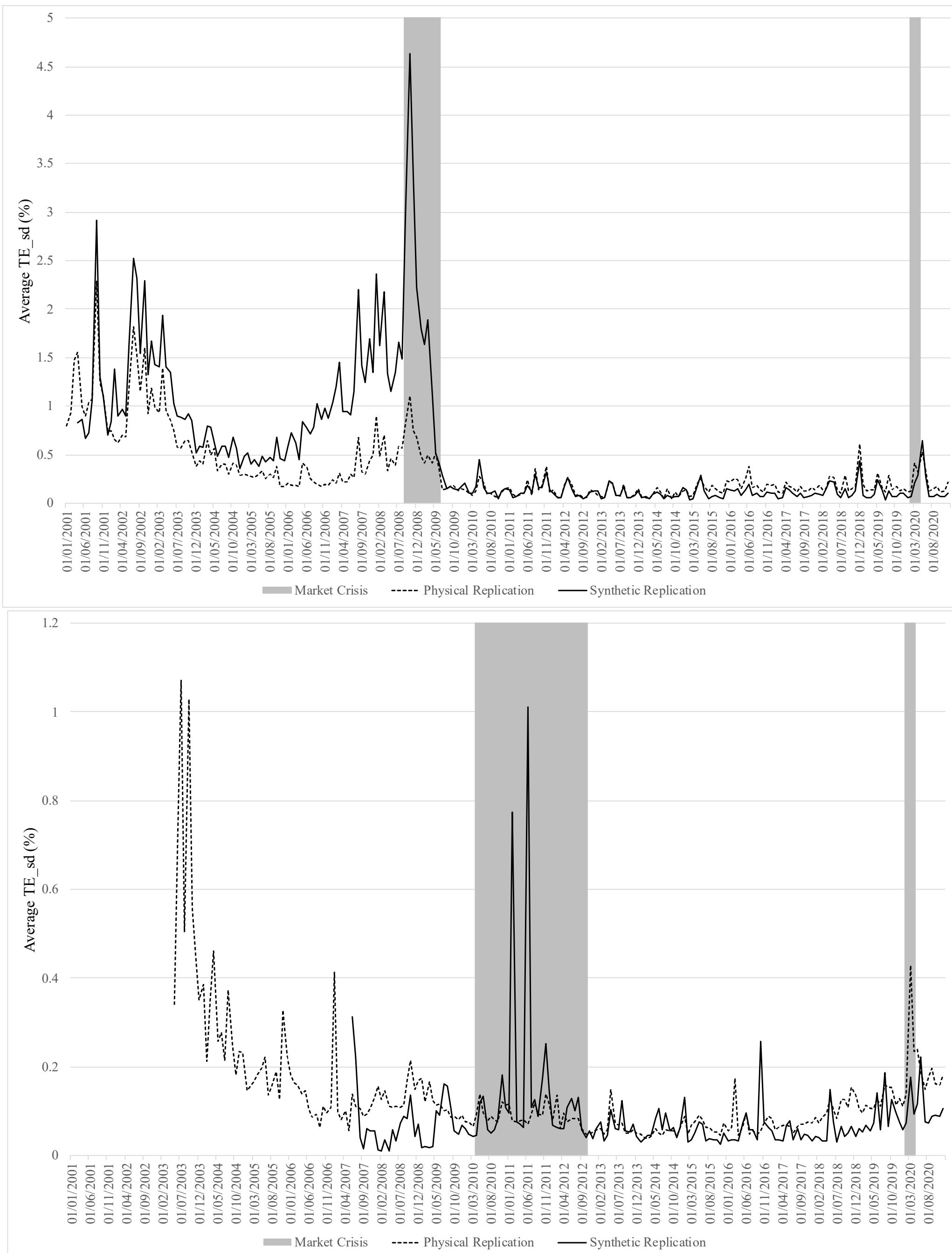


Figure 4.3. ANOVA Mean Plots

This figure plots the sample means with 95% confidence intervals from repeated measure ANOVA analysis. The whole sample is divided into four mutually exclusive groups. *Equity Pre*(*Post*) denotes the equity fund observations before (after) the end of the global financial crisis on June 2009, *FI Pre*(*Post*) denotes the fixed income fund observations before (after) the end of the sovereign debt crisis on October 2012. A different measure for tracking error is applied in each panel. TE_abs denotes the daily absolute tracking difference. TE_sd denotes the monthly standard deviation of daily excess returns over the benchmark. TE_sd_abs denotes the monthly standard deviation of daily absolute tracking differences. $TE\_se_r es$ denotes the standard error of residuals from the regression of daily fund returns on benchmark returns.

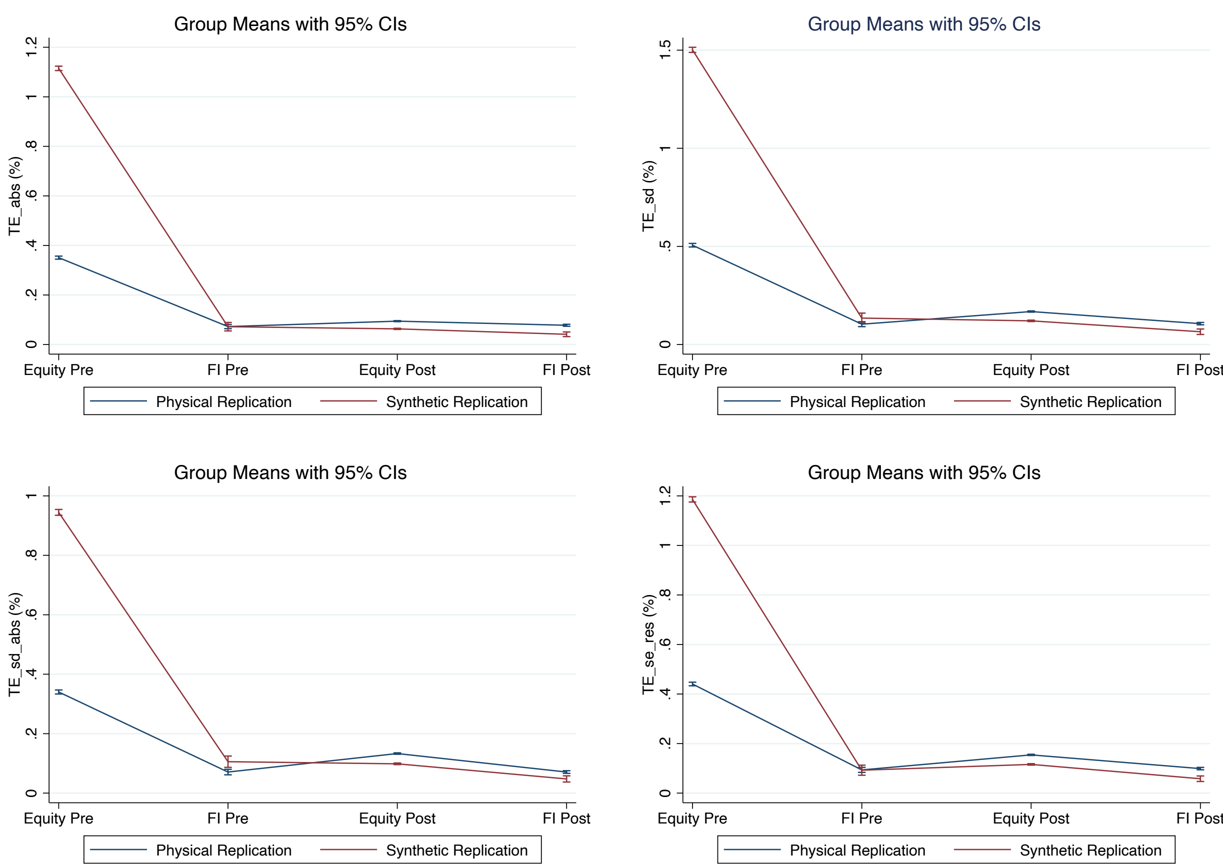
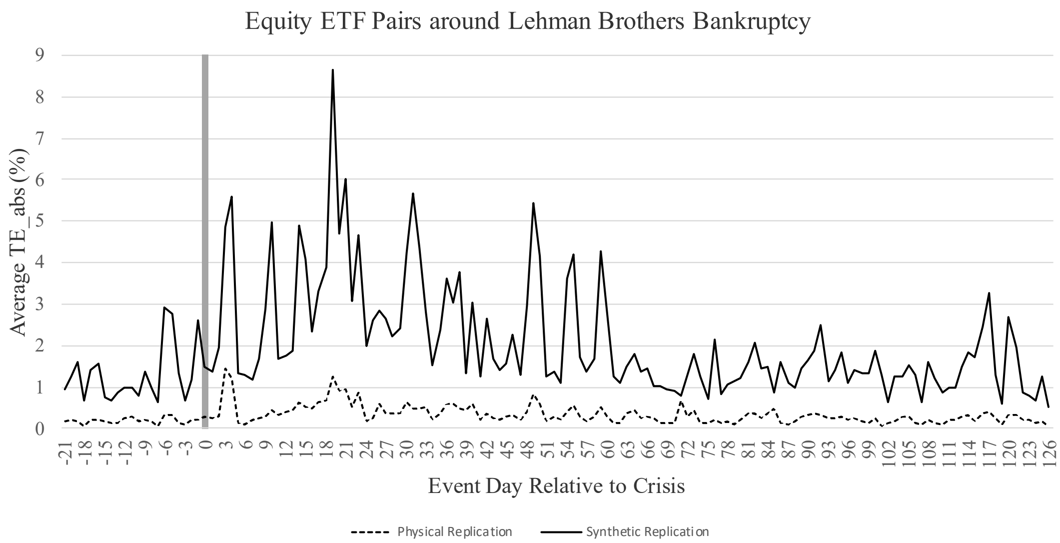
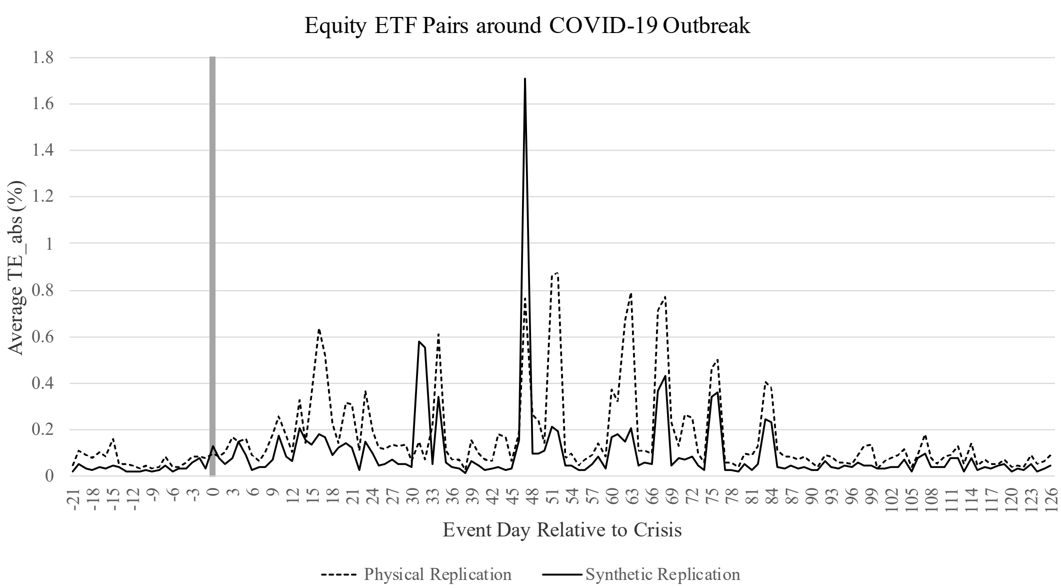
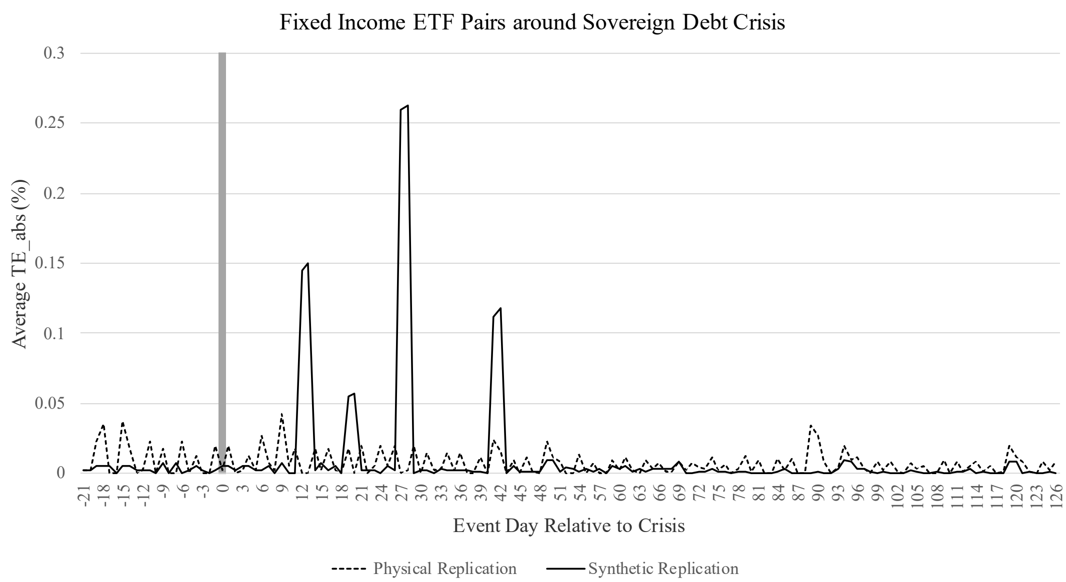


Figure 4.4. Event-time Daily Tracking Errors Between ETF Pairs

This figure displays the average daily absolute tracking differences for physical and synthetic ETF pairs following the same underlying benchmark in the [-21,+126] trading days event window around major market crises. The first two panels plot the equity ETF pairs around Lehman Brothers bankruptcy and COVID-19 outbreak. The last two panels plot the fixed income ETF pairs around sovereign debt crisis and COVID-19 outbreak.






Fixed Income ETF Pairs around Sovereign Debt Crisis
Average TE_abs (%)
Event Day Relative to Crisis
Physical Replication
Synthetic Replication

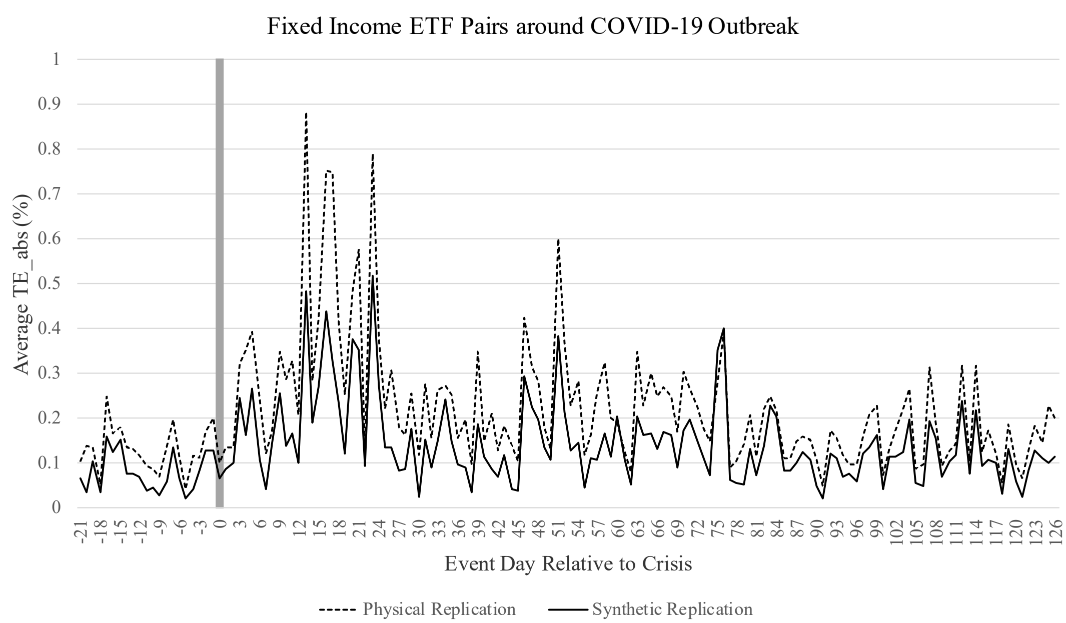

Fixed Income ETF Pairs around COVID-19 Outbreak
Average TE_abs (%)
Event Day Relative to Crisis
Physical Replication
Synthetic Replication

Table 4.1

Objective Distribution of Synthetic ETFs

This table reports the category distribution of synthetic ETFs. Panel A shows the 13 equity categories classified by geographic orientation and investment style. Panel B shows the 7 fixed income categories classified by geographic orientation. The last three columns compare the tracking efficiencies between two replication methods within each individual investment objective. The"MeanDiff" column reports the difference in the sample means, together with the statistical significance from a two-sided t-test. The asterisks denote statistical significance as follows: *** significant at 0.1%, ** significant at 1%, and * significant at 5%.

Category	Number of Funds	Number of Synthetic Funds	TE_sd (%) Physical	Synthetic	Diff.
Panel A: Global Equity Categories					
Sector Equity	402 (22%)	240 (60%)	0.249 (0.005)	0.207 (0.004)	0.042***
Europe Equity Large Cap	371 (20%)	78 (21%)	0.175 (0.003)	0.143 (0.006)	0.032***
APAC Equity	301 (17%)	128 (43%)	0.251 (0.003)	0.261 (0.006)	-0.010
US Equity Large Cap	248 (14%)	94 (38%)	0.236 (0.004)	0.166 (0.005)	0.070***
Global Equity	145 (8%)	41 (28%)	0.170 (0.006)	0.128 (0.006)	0.042***
Emerging Markets Equity	116 (6%)	58 (50%)	0.259 (0.012)	0.362 (0.012)	-0.103***
Europe Equity Mid/Small Cap	64 (4%)	18 (28%)	0.110 (0.004)	0.264 (0.019)	-0.154***
UK Equity Large Cap	51 (3%)	19 (37%)	0.104 (0.005)	0.161 (0.012)	-0.057***
Africa & Latin America Equity	39 (2%)	23 (59%)	0.275 (0.010)	0.459 (0.027)	-0.184***
Canada Equity	23 (1%)	4 (17%)	0.237 (0.008)	0.114 (0.011)	0.123***
US Equity Mid/Small Cap	23 (1%)	11 (48%)	0.194 (0.012)	0.178 (0.010)	0.016
Equity Miscellaneous	21 (1%)	6 (29%)	0.158 (0.013)	0.272 (0.018)	-0.114***
UK Equity Mid/Small Cap	11 (1%)	5 (45%)	0.039 (0.005)	0.182 (0.030)	-0.143***
All Equity	1508 (66%)	725 (48%)	0.204 (0.002)	0.219 (0.003)	-0.015***
Panel B: Global Fixed Income Categories					
Europe Fixed Income	195 (41%)	40 (21%)	0.062 (0.001)	0.068 (0.007)	-0.006
US Fixed Income	153 (32%)	3 (2%)	0.164 (0.003)	0.013 (0.002)	0.151***
Global Fixed Income	57 (12%)	5 (9%)	0.215 (0.004)	0.257 (0.013)	-0.042***
Emerging Markets Fixed Income	30 (6%)	0 (0%)			
UK Fixed Income	25 (5%)	5 (20%)	0.028 (0.002)	0.079 (0.024)	-0.051***
APAC Fixed Income	10 (2%)	1 (10%)	0.168 (0.007)	0.238 (0.016)	-0.070***
Fixed Income Miscellaneous	5 (1%)	2 (40%)	0.060 (0.014)	0.055 (0.010)	0.005
All Fixed Income	782 (34%)	56 (7%)	0.106 (0.001)	0.081 (0.006)	0.025***
All Funds	2290	781 (34%)	0.181 (0.001)	0.208 (0.002)	-0.027***

Table 4.2

Descriptive Statistics

This table reports the descriptive statistics for all four tracking error measures and major fund characteristics, including return volatility, benchmark alpha, TNA, age, expense ratio, bid-ask spread and dollar volume. Summary statistics on the daily change in levels of VIX, VSTOXX, NOISE and TED spread are also reported. In panel A, summary statistics are reported for the whole sample across the 20-year sample period 01 January 2001 to 31 December 2020. Panel B compares between equity and fixed income ETFs. Panel C compares between ETFs with physical replication and synthetic replication. The "MeanDiff" column in Panel B and C reports the difference in the sample means, together with the statistical significance from a two-sided t-test. The asterisks denote statistical significance as follows: *** significant at 0.1%, ** significant at 1%, and * significant at 5%.

Panel A: Descriptive Statistics for the Whole Sample

Variables	Mean	Std. Dev.	Median	Min	Max	Skew	Kurt
TE_abs (%)	0.117	0.514	0.010	0.000	98.690	21.335	1,420.981
TE_sd (%)	0.199	0.521	0.023	0.000	23.597	9.616	182.806
TE_sd_abs (%)	0.147	0.374	0.017	0.000	22.146	11.823	338.509
TE_se_res (%)	0.177	0.413	0.022	0.000	24.135	9.721	244.199
Return Volatility (%)	1.069	0.765	0.885	0.000	32.979	3.371	32.309
Benchmark Alpha (bps)	-0.078	36.489	-0.007	-390.718	371.096	-0.406	49.326
TNA ($million)	680	1,701	141	0.001	44,210	8.278	120.231
Age (years)	4.988	3.926	4.083	0.083	21.167	0.980	3.555
Expense Ratio (%)	0.367	0.191	0.330	0.000	1.810	0.884	4.626
Spread (%)	2.214	16.495	0.100	0.000	375.490	16.745	336.802
Dollar Volume ($million)	0.908	4.304	0.025	0.000	76.713	11.384	165.057
ΔVIX	-0.002	1.828	-0.090	-17.640	24.860	1.565	30.838
ΔVSTOXX	-0.001	1.853	-0.096	-13.987	22.642	1.458	22.270
ΔNOISE (bps)	0.000	0.322	0.000	-7.858	9.400	2.074	223.152
ΔTED (%)	0.000	0.050	0.000	-0.800	0.990	0.819	86.655

Panel B: Compare Between Equity and Fixed Income ETFs

Variables	Fixed Income		Equity		MeanDiff
	Obs.	Mean1	Obs.	Mean2	
TE_abs (%)	664,795	0.072	3,174,318	0.126	-0.053***
TE_sd (%)	780,699	0.203	2,954,322	0.198	0.005***
TE_sd_abs (%)	780,699	0.150	2,954,322	0.146	0.004***
TE_se_res (%)	780,051	0.178	2,949,232	0.177	0.001**
Return Volatility (%)	673,973	0.500	3,205,160	1.189	-0.689***
Benchmark Alpha (bps)	663,335	-0.073	3,168,835	-0.079	0.005
TNA ($million)	891,708	807	3,633,254	649	158***
Age (years)	643,511	4.333	3,219,436	5.119	-0.786***
Expense Ratio (%)	444,822	0.202	1,878,791	0.406	-0.204***
Spread (%)	866,348	2.197	2,907,915	2.219	-0.022
Dollar Volume ($million)	1,142,737	0.813	3,865,287	0.936	-0.123***

Panel C: Compare Between Physical and Synthetic ETFs

Variables	Physical		Synthetic		MeanDiff
	Obs.	Mean1	Obs.	Mean2	
TE_abs (%)	2,389,408	0.109	1,449,705	0.128	-0.019***
TE_sd (%)	2,314,358	0.195	1,420,663	0.206	-0.011***
TE_sd_abs (%)	2,314,358	0.145	1,420,663	0.149	-0.004***
TE_se_res (%)	2,309,729	0.175	1,419,554	0.181	-0.006***
Return Volatility (%)	2,409,877	1.017	1,469,256	1.155	-0.138***
Benchmark Alpha (bps)	2,384,832	-0.028	1,447,338	-0.160	0.132***
TNA ($million)	3,046,837	837	1,478,125	355	482***
Age (years)	2,443,566	5.112	1,419,381	4.776	0.336***
Expense Ratio (%)	1,738,485	0.353	585,128	0.409	-0.056***
Spread (%)	2,634,179	2.101	1,140,084	2.474	-0.374***
Dollar Volume ($million)	3,572,032	0.901	1,435,992	0.924	-0.023***

Table 4.3

Factors Affecting Tracking Errors

This table explores the factors affecting ETF tracking efficiencies through a pooled panel regression with AR (1) disturbance. The dependent variable is the daily absolute tracking difference for fund i on day t. *Synthetic* (*Equity*) is an indicator variable taking a value of one if fund i is a synthetic (equity) ETF. *Post Crisis* indicates the period after June 2009 in the equity subsample and after October 2012 in the fixed income subsample. $Return\ Volatility$ is the standard deviation of daily excess returns calculated from a rolling window of 21 trading days. *BenchmarkAlpha* is the difference between the daily realized return and the predicted return with beta estimated from the single index model using a rolling window of 252 trading days. Time fixed effect is included in the full sample, category fixed effect is included in the equity and fixed income subsamples. All variables are standardized to have a mean of zero and a standard deviation of one. *Rho_AR* is the autocorrelation coefficient estimated in error terms from Prais and Winsten (1954). ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively.

Variables	$TE\_abs_{i,t}$		
	All Funds	Equity	Fixed Income
Synthetic*Equity	0.059		
	(0.070)		
Equity	0.001		
	(0.026)		
Synthetic*Post Crisis		-2.488***	0.020
		(0.034)	(0.033)
Post Crisis		-0.289***	0.034**
		(0.014)	(0.017)
Synthetic	-0.054	2.352***	-0.034
	(0.067)	(0.039)	(0.083)
Return Volatility$_{i,t-1}$	0.034***	0.069***	0.133***
	(0.004)	(0.003)	(0.004)
Benchmark Alpha$_{i,t-1}$	-0.021***	-0.021***	-0.026***
	(0.001)	(0.001)	(0.002)
Log (TNA)$_{i,t-1}$	0.021***	0.007	0.025***
	(0.006)	(0.007)	(0.009)
Age$_{i,t-1}$	-0.045***	-0.022***	0.008
	(0.008)	(0.005)	(0.008)
Expense Ratio$_{i,t-1}$	0.014*	-0.004	0.050***
	(0.008)	(0.009)	(0.017)
Bid-Ask Spread$_{i,t-1}$	0.001	0.001	0.000
	(0.001)	(0.001)	(0.002)
Dollar Volume$_{i,t-1}$	-0.000	-0.000	-0.002
	(0.001)	(0.001)	(0.002)
Constant	0.224	0.328***	0.172
	(0.255)	(0.031)	(0.148)
Observations	445,930	363,780	82,150
Number of fundid	1,295	1,026	269
Time FE	Yes	No	No
Category FE	No	Yes	Yes
Adjusted R^2	0.135	0.136	0.135
Rho_AR	0.652	0.659	0.501

Table 4.4

Difference-in-Difference Analysis around Market Crisis

This table reports the results from Difference-in-Difference (DiD) analysis on relative tracking efficiencies between ETF pairs following the same underlying benchmarks. The dependent variable is the daily absolute tracking difference for fund i on event day t. *Event Day* represents 15 September 2008 for Lehman Brothers bankruptcy, 27 April 2010 for sovereign debt crisis and 20 Febuary 2020 for COVID-19 outbreak. Event window $[+1, +5]$ denotes the first week post event, $[+6, +21]$ denotes the rest of the first month and $[+22, +126]$ denotes up until half a year post event. Panel A compares how tracking performance of equity ETF pairs reacts differently towards Lehman Brothers bankruptcy and COVID-19 outbreak. Panel B compares fixed income ETF pairs around sovereign debt crisis and COVID-19. Category fixed effect is included in all model specifications. All variables are standardized to have a mean of zero and a standard deviation of one. *Rho_AR* is the autocorrelation coefficient estimated in error terms from Prais and Winsten (1954). ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively.

Panel A: Equity ETF Pairs

Variables	$TE\_abs_{i,t}$			
	Lehman Brothers		COVID-19	
	(1)	(2)	(3)	(4)
Synthetic*Event Day	-0.043	0.009	0.173**	0.009
	(0.124)	(0.404)	(0.083)	(0.241)
Synthetic*[+1, +5]	0.581***	0.484*	-0.012	-0.008
	(0.077)	(0.279)	(0.058)	(0.185)
Synthetic*[+6, +21]	0.785***	0.811***	-0.186***	-0.305**
	(0.055)	(0.214)	(0.042)	(0.148)
Synthetic*[+22, +126]	0.250***	0.171	-0.079***	-0.049
	(0.041)	(0.168)	(0.030)	(0.112)
Synthetic	0.304***	0.578***	-0.060	-0.178
	(0.088)	(0.196)	(0.044)	(0.114)
Event Day	0.049	0.202	0.066	0.157
	(0.093)	(0.350)	(0.055)	(0.157)
[+1, +5]	0.081	0.400*	0.168***	0.122
	(0.058)	(0.230)	(0.039)	(0.113)
[+6, +21]	0.106**	0.167	0.381***	0.288***
	(0.041)	(0.184)	(0.028)	(0.092)
[+22, +126]	0.034	0.031	0.231***	0.118*
	(0.031)	(0.145)	(0.020)	(0.070)
Return Volatility$_{i,t-1}$		0.072**		0.102***
		(0.033)		(0.020)
Benchmark Alpha$_{i,t-1}$		-0.028**		-0.150***
		(0.013)		(0.007)
Log (TNA)$_{i,t-1}$		0.198***		0.017
		(0.075)		(0.040)
Age$_{i,t-1}$		-0.316***		-0.116***
		(0.075)		(0.028)
Expense Ratio$_{i,t-1}$		0.390***		-0.041
		(0.108)		(0.038)
Bid-Ask Spread$_{i,t-1}$		0.018		0.013
		(0.015)		(0.009)
Dollar Volume$_{i,t-1}$		0.028*		-0.013
		(0.015)		(0.010)
Constant	-0.278*	-0.492*	-0.067	0.015
	(0.168)	(0.298)	(0.047)	(0.106)
Observations	23,088	2,249	63,660	10,458
Number of fundid	179	65	441	165
Category FE	Yes	Yes	Yes	Yes
Adjusted R^2	0.158	0.329	0.026	0.097
Rho_AR	0.259	0.518	0.358	0.604

Panel B: Fixed Income ETF Pairs		
Variables	TE_abs$_{i,t}$	
	Sovereign Debt Crisis	COVID-19
Synthetic*Event Day	0.700	0.045
	(0.621)	(0.317)
Synthetic*[+1, +5]	0.352	-0.145
	(0.467)	(0.181)
Synthetic*[+6, +21]	0.658*	-0.398***
	(0.344)	(0.124)
Synthetic*[+22, +126]	0.290	-0.061
	(0.246)	(0.089)
Synthetic	-0.248	0.028
	(0.230)	(0.206)
Event Day	-0.588	-0.062
	(0.439)	(0.186)
[+1, +5]	-0.182	0.474***
	(0.330)	(0.106)
[+6, +21]	0.128	0.906***
	(0.243)	(0.073)
[+22, +126]	-0.073	0.230***
	(0.177)	(0.053)
Constant	0.072	-0.846***
	(0.163)	(0.193)
Observations	1,613	3,858
Number of fundid	13	27
Category FE	Yes	Yes
Adjusted R^2	0.027	0.203
Rho_AR	0.419	0.171

Table 4.5

Principal Component Analysis

This table presents the results from principle component analysis used to extract the common factor from state variables. VIX is the CBOE volatility index, which measures the expectation of stock market volatility over the coming 30 days. $VSTOXX$ is the Euro Stoxx 50 volatility index, commonly known as the "Euro VIX", which measures the implied volatility of near term EuroStoxx 50 options. $NOISE$ is a market-wide liquidity measure proposed by Hu et al. (2013). TED is calculated as the spread between 3-month LIBOR based on US dollars and 3-month Treasury Bill. Panel A lists the eigenvalues and proportion of variance explained by the four principle components using 4,860 days when all state variables are jointly available. Panel B displays the eigenvectors, i.e. factor loadings, on the principle components. Panel C shows the correlation matrix between the estimated first principle component ($PC1$) and the four state variables.

Panel A: Eigenvalues and Proportion of Variance Explained by Principal Components (N = 4,860 days)

Principal Component (PC)	Eigenvalue	Difference	% Variance Explained	Cumulative % Variance
1	2.784	1.958	69.6%	69.6%
2	0.827	0.509	20.7%	90.3%
3	0.317	0.246	7.9%	98.2%
4	0.0716	.	1.8%	100.0%

Panel B: Eigenvectors (Factor Loadings) on Principal Components

Variable	PC1	PC2	PC3	PC4
VIX	0.558	-0.324	0.139	-0.751
VSTOXX	0.512	-0.524	0.216	0.646
NOISE	0.510	0.294	-0.802	0.103
TED	0.409	0.731	0.539	0.088

Panel C: Correlation Between PC1 and State Variables

	PC1	VIX	VSTOXX	NOISE	TED
PC1	1				
VIX	0.931	1			
VSTOXX	0.854	0.910	1		
NOISE	0.851	0.671	0.548	1	
TED	0.682	0.458	0.307	0.622	1

Table 4.6
Tracking Performance Sensitivity to Market Distress - Equity Subsample

This table explores how tracking performance sensitivity to market distress for equity ETFs varies across time, using a pooled panel regression with AR (1) disturbance. The dependent variable is the daily absolute tracking difference for fund i on day t. *Synthetic* is an indicator variable taking a value of one if fund i is a synthetic ETF. *Post Crisis* indicates the period after June 2009. *Recession* marks the NBER defined US market recession periods. ΔVIX ($\Delta VSTOXX$) is the daily change in CBOE (Euro Stoxx 50) volatility index. $\Delta PC1$ is the daily change in the common factor extracted from all four state variables using PCA. All continuous explanatory variables in Table 4.3 are included as controls. Category fixed effect is included. All variables are standardized to have a mean of zero and a standard deviation of one. *Rho_AR* is the autocorrelation coefficient estimated in error terms from Prais and Winsten (1954). ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively.

Variables	$TE\_abs_{i,t}$				
	(1)	(2)	(3)	(4)	(5)
Synthetic*Post Crisis*Recession	-1.718***				-1.711***
	(0.126)				(0.131)
Synthetic*Recession	1.634***				1.616***
	(0.121)				(0.125)
Post Crisis*Recession	-0.352***				-0.356***
	(0.028)				(0.029)
Recession	0.280***				0.292***
	(0.023)				(0.023)
Synthetic*Post Crisis*$\Delta VIX_{i,t-1}$		-0.036***			
		(0.006)			
Synthetic*$\Delta VIX_{i,t-1}$		0.043***			
		(0.006)			
Post Crisis*$\Delta VIX_{i,t-1}$		-0.003			
		(0.003)			
$\Delta VIX_{i,t-1}$		0.005			
		(0.003)			
Synthetic*Post Crisis*$\Delta VSTOXX_{i,t-1}$			-0.094***		
			(0.007)		
Synthetic*$\Delta VSTOXX_{i,t-1}$			0.104***		
			(0.006)		
Post Crisis*$\Delta VSTOXX_{i,t-1}$			-0.070***		
			(0.004)		
$\Delta VSTOXX_{i,t-1}$			0.069***		
			(0.003)		
Synthetic*Post Crisis*$\Delta PC1_{i,t-1}$				-0.029***	-0.028***
				(0.007)	(0.007)
Synthetic*$\Delta PC1_{i,t-1}$				0.039***	0.037***
				(0.006)	(0.006)
Post Crisis*$\Delta PC1_{i,t-1}$				-0.013***	-0.014***
				(0.004)	(0.004)
$\Delta PC1_{i,t-1}$				0.016***	0.017***
				(0.003)	(0.003)
Synthetic*Post Crisis	-0.840***	-2.500***	-2.482***	-2.469***	-0.832***
	(0.117)	(0.034)	(0.034)	(0.035)	(0.121)
Synthetic	0.709***	2.365***	2.346***	2.337***	0.706***
	(0.119)	(0.039)	(0.039)	(0.040)	(0.123)
Post Crisis	-0.130***	-0.288***	-0.289***	-0.290***	-0.123***
	(0.019)	(0.014)	(0.014)	(0.015)	(0.020)
Constant	0.168***	0.327***	0.328***	0.329***	0.161***
	(0.033)	(0.031)	(0.031)	(0.031)	(0.034)
Observations	363,780	355,530	363,780	333,255	333,255
Number of fundid	1,026	1,026	1,026	1,026	1,026
Controls & Category FE	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.141	0.137	0.138	0.136	0.141
Rho_AR	0.658	0.663	0.661	0.675	0.674

Table 4.7

Tracking Performance Sensitivity to Market Distress - Fixed Income Subsample

This table explores how tracking performance sensitivity to market distress for fixed income ETFs varies across time, using a pooled panel regression with AR (1) disturbance. The dependent variable is the daily absolute tracking difference for fund i on day t. *Synthetic* is an indicator variable taking a value of one if fund i is a synthetic ETF. *Post Crisis* indicates the period after October 2012. ΔVIX ($\Delta VSTOXX$) is the daily change in CBOE (Euro Stoxx 50) volatility index. $\Delta NOISE$ is the daily change in the liquidity measure proposed by (Hu et al., 2013). ΔTED is the daily change in TED Spread. $\Delta PC1$ is the daily change in the common factor extracted from all four state variables using PCA. All continuous explanatory variables in Table 4.3 are included as controls. Category fixed effect is included. All variables are standardized to have a mean of zero and a standard deviation of one. *Rho_AR* is the autocorrelation coefficient estimated in error terms from Prais and Winsten (1954). ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively.

Variables	$TE\_abs_{i,t}$				
	(1)	(2)	(3)	(4)	(5)
Synthetic*Post Crisis*$\Delta VIX_{i,t-1}$	-0.004				
	(0.013)				
Synthetic*$\Delta VIX_{i,t-1}$	0.015				
	(0.011)				
Post Crisis*$\Delta VIX_{i,t-1}$	-0.008*				
	(0.005)				
$\Delta VIX_{i,t-1}$	0.003				
	(0.005)				
Synthetic*Post Crisis*$\Delta VSTOXX_{i,t-1}$		-0.028**			
		(0.014)			
Synthetic*$\Delta VSTOXX_{i,t-1}$		0.044***			
		(0.012)			
Post Crisis*$\Delta VSTOXX_{i,t-1}$		-0.008			
		(0.005)			
$\Delta VSTOXX_{i,t-1}$		0.002			
		(0.005)			
Synthetic*Post Crisis*$\Delta Noise_{i,t-1}$			-0.005		
			(0.017)		
Synthetic*$\Delta Noise_{i,t-1}$			0.001		
			(0.014)		
Post Crisis*$\Delta Noise_{i,t-1}$			-0.001		
			(0.007)		
$\Delta Noise_{i,t-1}$			0.004		
			(0.006)		
Synthetic*Post Crisis*$\Delta TED_{i,t-1}$				-0.051	
				(0.053)	
Synthetic*$\Delta TED_{i,t-1}$				0.022	
				(0.051)	
Post Crisis*$\Delta TED_{i,t-1}$				0.022***	
				(0.008)	
$\Delta TED_{i,t-1}$				0.002	
				(0.006)	
Synthetic*Post Crisis*$\Delta PC1_{i,t-1}$					-0.030**
					(0.015)
Synthetic*$\Delta PC1_{i,t-1}$					0.041***
					(0.013)
Post Crisis*$\Delta PC1_{i,t-1}$					-0.008
					(0.006)
$\Delta PC1_{i,t-1}$					0.005
					(0.005)
Synthetic*Post Crisis	0.020	0.020	0.020	0.016	0.019
	(0.034)	(0.033)	(0.035)	(0.035)	(0.036)
Synthetic	-0.030	-0.034	-0.031	-0.023	-0.025
	(0.083)	(0.083)	(0.083)	(0.083)	(0.083)
Post Crisis	0.035**	0.034**	0.035**	0.038**	0.042**
	(0.017)	(0.017)	(0.018)	(0.018)	(0.018)
Constant	0.162	0.172	0.163	0.157	0.163
	(0.148)	(0.148)	(0.148)	(0.148)	(0.149)
Observations	80,281	82,150	79,729	78,861	75,377
Number of fundid	269	269	269	269	269
Controls & Category FE	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.134	0.135	0.133	0.133	0.135
Rho_AR	0.518	0.502	0.522	0.525	0.532

Table A4.1
Top 20 European ETF Providers

This table list the top 20 European ETF providers ranked by the aggregate family TNA by the end of 2020. Family TNA is reported in billion USD. The total number of ETF offerings as well as the number and percentage of fund offerings with synthetic replication are reported. The percentage of family assets under synthetic structure is reported in the last column.

Rank (by TNA)	Branding Name	Family TNA ($billion)	Number of ETFs	Number of Synthetic ETFs	% Number of Synthetic ETFs	% TNA of Synthetic ETFs
1	iShares	1390.573	587	3	0.51%	0.02%
2	UBS	447.717	424	37	8.73%	22.69%
3	Xtrackers	322.632	342	136	39.77%	23.57%
4	Lyxor	280.834	651	427	65.59%	51.26%
5	Amundi	210.466	350	223	63.71%	47.80%
6	Vanguard	144.023	52	0	0.00%	0.00%
7	Invesco	99.223	179	93	51.96%	59.18%
8	State Street	80.753	138	0	0.00%	0.00%
9	BNP Paribas	49.888	130	57	43.85%	45.68%
10	HSBC	11.790	37	0	0.00%	0.00%
11	Deka	11.589	54	1	1.85%	0.52%
12	JPMorgan	10.180	29	5	17.24%	0.00%
13	PIMCO	9.815	10	0	0.00%	0.00%
14	Natixis	7.741	23	11	47.83%	87.12%
15	Legal & General	7.673	33	4	12.12%	3.49%
16	Credit Suisse	7.540	10	0	0.00%	0.00%
17	Handelsbanken	4.426	18	12	66.67%	4.99%
18	Fidelity	4.071	13	0	0.00%	0.00%
19	VanEck	3.631	23	0	0.00%	0.00%
20	WisdomTree	2.991	56	0	0.00%	0.00%

Table A4.2

Variable Definitions

Variable	Definition
	Panel A: Tracking Error Measures
TE_abs	Daily absolute difference between the fund return and the benchmark return.
TE_sd	Standard deviation of fund daily excess returns over the benchmark.
TE_sd_abs	Standard deviation of daily absolute tracking difference.
TE_se_res	Standard Error of residuals from the following regression: $R_{i,t} = \alpha_i + \beta_{i,t} R_{BMK_i,t} + \varepsilon_{i,t}$.
	Panel B: Fund Characteristics
Return Volatility	Standard deviation of excess returns over the risk-free rate for the prior 21 trading days.
Benchmark Alpha	Difference between the daily realized return and the predicted return with beta estimated from the single index model using a rolling window of 252 trading days.
TNA	Total Net Asset managed by the fund obtained from Morningstar.
Age	Number of years since the inception of fund.
Expense Ratio	Annual report total expense ratio obtained from Morningstar. Including all asset-based cost incurred by the fund, expressed in percentage of assets.
Bid-Ask Spread	Daily Bid-Ask spread obtained from Morningstar.
Dollar Volume	Daily trading volume multiplied by daily closing price, both obtained from Morningstar.
	Panel C: State Variables
VIX	CBOE volatility index (VIX) obtained from FRED, which measures the expectation of stock market volatility over the coming 30 days.
VSTOXX	Euro Stoxx 50 volatility index obtained from Qontigo, which measures the implied volatility of near term EuroStoxx 50 options.
NOISE	A market-wide liquidity measure proposed by Hu et al. (2013).
TED	TED spread is calculated as the spread between 3-month LIBOR based on US dollars and 3-month Treasury Bill.
PC1	The first principle component extracted from the above four state variables from principal component analysis.
Recession	An indicator variable taking a a value of one during three NBER defined market recession periods, namely March 2001 to November 2001, December 2007 to June 2009 and Febuary 2020 to April 2020.
Post Crisis	An indicator variable that marks the period June 2009 to December 2020 in the equity subsample, and October 2012 to December 2020 in the fixed income subsample.

Chapter 5

Conclusion

This thesis studies three related topics in asset management, namely the determinants of ETF launching decisions, the impact of MF subadvisor turnovers, and the effect of replication method on ETF tracking efficiencies.

The first essay contributes to the industrial organization literature on the ETF industry. First, I build upon the literature concerning the industrial organization of the open-ended funds to show how the decision to launch an ETF is affected by fund characteristics in distinct ways relative to open-ended MFs. Second, I provide evidence on how market conditions, including liquidity and market concentration, affect the competition and growth of the ETF industry. Prior literature has documented the effect of market quality on ETF flows (Clifford et al., 2014). In this essay, I discuss a more direct channel of the industry expansion, namely the emergence of new market entrants.

The second essay makes three main contributions. First, we analyse what drives the large number of hiring and firing events and their resulting impact on mutual funds and sub-advisors' returns. Second, we study competition among sub-advisors by exploiting the fact that the existence of sub-advised (or outsourced) funds with two or more sub-advisors effectively introduces competition into the management of these funds. The presence of more than one sub-advisor leads to a split of management fees among sub-advisors, which in turn encourages greater monitoring between them and appears to benefit performance (Moreno et al. (2018)). Exploiting the fact that 32% of the outsourced funds in our sample are managed by two or more sub-advisors, we examine if the performance of the incumbent sub-advisors varies when either one of them is fired or a new sub-advisor is hired and if this performance is different from that of the recently hired/fired sub-advisors. Finally, we study how actively the fund family monitors its sub-advisors beyond looking at performance and whether deviating from the investment mandate affects a subadvisor's chances of being hired or fired. Rather than outsourcing asset management to improve returns, advisors just ensure that the mandates given to sub-advisors are followed.

The third essay contributes to the literature from three perspectives. First, I disentangle the effect of market distress on ETF tracking efficiencies. Difference in tracking performance deterioration around market crisis between ETFs with physical and synthetic replication is attributed to their different reactions to counterparty risk and liquidity risk. Second, the extensive 20-year sample period allows identification of several physical and synthetic ETF pairs tracking the same underlying benchmarks, and therefore a like-for-like comparison across different types of major market crisis. To my best knowledge, this is the first study to make a direct comparison of ETFs with different replication methods on the same underlying index.[1] Third, I identify a structural break in both the level and sensitivity of market-wide tracking efficiencies after the global financial crisis, and provide empirical evidence on how market failure evokes tighter risk management.

Finally, I list here several directions for future research. Regarding the first essay, it is interesting to examine the family decision to launch an ETF versus an index fund. There are many similarities shared between ETFs and index funds, including their passive and low-cost nature, as well as their tax efficiencies. While I show that ETF launching decisions are affected by different factors compared to the active funds, the difference between launching an ETF and an index fund could be more subtle. Regarding the second essay, we plan to further collect data on the distribution channel and in-house managing capacity of fund families, to see if they explain cross-sectional variations in the turnover effects. Regarding the third essay, there are three directions to extend the research. First, subject to the accessibility of information regarding the swap counterparties for synthetic ETFs, one could study the relation between

[1] Meinhardt et al. (2014) explains the scarcity in physical and synthetic ETFs mimicking the same index.

tracking efficiency and the level of credit risk in the swap contract. Second, an event study on the different reactions of tracking efficiencies could be carried out around dividend payments by top constituents of the underlying index. Third, it worth checking if there are cases where ETFs with synthetic structure before the global financial crisis are transformed into physical structure afterwards, and if these contribute to the post-crisis reversion in relative tracking efficiencies.

Bibliography

Antoniewicz, R. S. and Heinrichs, J. (2014). Understanding Exchange-Traded Funds: How ETFs Work. SSRN Scholarly Paper ID 2523540, Social Science Research Network, Rochester, NY.

Ben-David, I., Franzoni, F. A., Kim, B., and Moussawi, R. (2021). Competition for Attention in the ETF Space. SSRN Scholarly Paper ID 3765063, Social Science Research Network, Rochester, NY.

Ben-David, I., Franzoni, F., and Moussawi, R. (2018). Do ETFs Increase Volatility? *The Journal of Finance*, 73(6):2471–2535. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12727.

Bergstresser, D. and Poterba, J. (2002). Do after-tax returns affect mutual fund inflows? *Journal of Financial Economics*, 63(3):381–414.

Berk, J. and Green, R. (2004). Mutual Fund Flows and Performance in Rational Markets. *Journal of Political Economy*, 112(6):1269–1295. Publisher: The University of Chicago Press.

BIS (2011). The impact of sovereign credit risk on bank funding conditions. *Bank for International Settlements*.

Brown, S. J. and Goetzmann, W. N. (1995). Performance Persistence. *The Journal of Finance*, 50(2):679–698. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1995.tb04800.x.

Buetow, G. and Henderson, B. (2012). An Empirical Analysis of Exchange-Traded Funds. *Journal of Portfolio Management*, 38:112–127.

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal* *of Finance*, 52(1):57–82. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1997.tb03808.x.

Chen, J., Hong, H., Jiang, W., and Kubik, J. D. (2013). Outsourcing Mutual Fund Management: Firm Boundaries, Incentives, and Performance. *The Journal of Finance*, 68(2):523–558. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12006.

Chevalier, J. and Ellison, G. (1997). Risk Taking by Mutual Funds as a Response to Incentives. *Journal of Political Economy*, 105(6):1167–1200. Publisher: The University of Chicago Press.

Chevalier, J. and Ellison, G. (1999). Career Concerns of Mutual Fund Managers*. *The Quarterly Journal of Economics*, 114(2):389–432.

Chu, P. K.-K. (2011). Study on the tracking errors and their determinants: evidence from Hong Kong exchange traded funds. *Applied Financial Economics*, 21(5):309–315. Publisher: Routledge _eprint: https://doi.org/10.1080/09603107.2010.530215.

Chuprinin, O., Massa, M., and Schumacher, D. (2015). Outsourcing in the International Mutual Fund Industry: An Equilibrium View. *The Journal of Finance*, 70(5):2275–2308. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12259.

Clifford, C. P., Fulkerson, J. A., and Jordan, B. D. (2014). What Drives ETF Flows? *Financial Review*, 49(3):619–642.

Daniel, K., Grinblatt, M., Titman, S., and Wermers, R. (1997). Measuring Mutual Fund Performance with Characteristic-Based Benchmarks. *The Journal of Finance*, 52(3):1035–1058. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1997.tb02724.x.

Debaere, P. M. and Evans, R. B. (2014). Outsourcing vs. Integration in the Mutual Fund Industry: An Incomplete Contracting Perspective. SSRN Scholarly Paper ID 2399177, Social Science Research Network, Rochester, NY.

ECB (2020). COVID-19 and the liquidity crisis of non-banks: lessons for the future. *European Central Bank*.

Elia, M. (2011). Tracking Error of Traditional and Synthetic European Exchange-Traded Funds. page 22.

Elton, E. J., Gruber, M. J., and Blake, C. R. (1996). The Persistence of Risk-Adjusted

Mutual Fund Performance. *The Journal of Business*, 69(2):133–157. Publisher: University of Chicago Press.

Elton, E. J., Gruber, M. J., and Blake, C. R. (2003). Incentive Fees and Mutual Funds. *The Journal of Finance*, 58(2):779–804. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1540-6261.00545.

ESMA (2012). A ESMA's Guidelines on ETFs and Other UCITS Issues. *European Securities and Markets Authority Consultation Paper*.

Evans, R. B. (2010). Mutual Fund Incubation. *The Journal of Finance*, 65(4):1581–1611.

Fender, I. and Gyntelberg, J. (2008). Three market implications of the Lehman bankruptcy. *BIS Quarterly Review*.

Frino, A. and Gallagher, D. R. (2001). Tracking S&P 500 Index Funds. *The Journal of Portfolio Management*, 28(1):44–55. Publisher: Institutional Investor Journals Umbrella Section: Primary Article.

Garleanu, N. and Pedersen, L. H. (2019). Active and Passive Investing. SSRN Scholarly Paper ID 3253537, Social Science Research Network, Rochester, NY.

Goyal, A. and Wahal, S. (2008). The Selection and Termination of Investment Management Firms by Plan Sponsors. *The Journal of Finance*, 63(4):1805–1847. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2008.01375.x.

Greene, J. T. and Stark, J. (2016). What's Trending? The Performance and Motivations for Mutual Fund Startups. SSRN Scholarly Paper ID 2826677, Social Science Research Network, Rochester, NY.

Grill, M., Lambert, C., Philipp, M., Watfe, G., and Weistroffer, C. (2018). Counterparty and Liquidity Risks in Exchange-traded Funds. *European Central Bank*.

Grinblatt, M. and Titman, S. (1992). The Persistence of Mutual Fund Performance. *The Journal of Finance*, 47(5):1977–1984. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1992.tb04692.x.

Guercio, D. D. and Tkac, P. A. (2002). The Determinants of the Flow of Funds of Managed Portfolios: Mutual Funds vs. Pension Funds. *Journal of Financial and Quantitative Analysis*, 37(4):523–557. Publisher: Cambridge University Press.

Hasbrouck, J. and Seppi, D. J. (2001). Common factors in prices, order flows, and liquidity. *Journal of Financial Economics*, 59(3):383–411.

Hu, G. X., Pan, J., and Wang, J. (2013). Noise as Information for Illiquidity. *The Journal of Finance*, 68(6):2341–2382. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12083.

Hurlin, C., Iseli, G., Pérignon, C., and Yeung, S. (2019). The counterparty risk exposure of ETF investors. *Journal of Banking & Finance*, 102:215–230.

Inoue, A. and Solon, G. (2006). A Portmanteau Test for Serially Corelated Errors in Fixed Effects Models. *Econometric Theory*, 22(5):835–851. Publisher: Cambridge University Press.

Ippolito, R. A. (1992). Consumer Reaction to Measures of Poor Quality: Evidence from the Mutual Fund Industry. *The Journal of Law and Economics*, 35(1):45–70. Publisher: The University of Chicago Press.

Johnson, B., Bioy, H., Kellett, A., and Davidson, L. (2013). On the Right Track: *Measuring Tracking Efficiency in ETFs*. *The Journal of Index Investing*, 4(3):35–41.

Johnson, T. C. (2008). Volume, liquidity, and liquidity risk. *Journal of Financial Economics*, 87(2):388–417.

Karceski, J. (2002). Returns-Chasing Behavior, Mutual Funds, and Beta's Death. *The Journal of Financial and Quantitative Analysis*, 37(4):559–594. Publisher: [Cambridge University Press, University of Washington School of Business Administration].

Khorana, A. and Servaes, H. (1999). The Determinants of Mutual Fund Starts. *The Review of Financial Studies*, 12(5):1043–1074.

Knittel, C. R., Heisler, J., Neumann, J. J., and Stewart, S. (2004). Why do institutional plan sponsors hire and fire their investment managers? Working Paper 05-27, Working Paper.

Korajczyk, R. A. and Sadka, R. (2008). Pricing the commonality across alternative measures of liquidity. *Journal of Financial Economics*, 87(1):45–72.

Kostovetsky, L. and Warner, J. B. (2015). You're Fired! New Evidence on Portfolio Manager Turnover and Performance. *Journal of Financial and Quantitative Analysis*, 50(4):729–755.

Publisher: Cambridge University Press.

Levy, A. and Lieberman, O. (2016). Active Flows and Passive Returns. *Review of Finance*, 20(1):373–401. Publisher: Oxford Academic.

Loughran, T. and Ritter, J. R. (2002). Why Don't Issuers Get Upset About Leaving Money on the Table in IPOs? *The Review of Financial Studies*, 15(2):32.

Lunde, A., Timmermann, A., and Blake, D. (1999). The hazards of mutual fund underperformance: A Cox regression analysis. *Journal of Empirical Finance*, 6(2):121–152.

Magkotsios, G. (2018). Industry-Level Returns to Scale and Investor Flows in Asset Management. SSRN Scholarly Paper ID 3206239, Social Science Research Network, Rochester, NY.

Mateus, C. and Rahmani, Y. (2017). Physical versus Synthetic Exchange Traded Funds. Which One Replicates Better? *Journal of Mathematical Finance*, 07:975–989.

Meinhardt, C., Mueller, S., and Schoene, S. (2014). Physical and Synthetic Exchange Traded Funds: the Good, the Bad or the Ugly? SSRN Scholarly Paper ID 2026409, Social Science Research Network, Rochester, NY.

Moreno, D., Rodríguez, R., and Zambrana, R. (2018). Management sub-advising in the mutual fund industry. *Journal of Financial Economics*, 127(3):567–587.

Naumenko, K. and Chystiakova, O. (2015). An Empirical Study on the Differences between Synthetic and Physical ETFs. *International Journal of Economics and Finance*, 7(3):p24. Number: 3.

Pope, P. F. and Yadav, P. K. (1994). Discovering Errors in Tracking Error. *The Journal of Portfolio Management*, 20(2):27–32.

Prais, S. J. and Winsten, C. B. (1954). Trend Estimators and Serial Correlation. *Cowles Commission Discussion Paper*, No. 383(Chicago).

Pástor, L., Stambaugh, R. F., and Taylor, L. A. (2015). Scale and skill in active management. *Journal of Financial Economics*, 116(1):23–45.

Ramaswamy, S. (2011). Market structures and systemic risks of exchange-traded funds. BIS Working Paper 343, Bank for International Settlements.

Reinganum, J. F. (1985). Innovation and Industry Evolution. *The Quarterly Journal of Economics*, 100(1):81–99. Publisher: Oxford Academic.

Richardson, S., Saffi, P. A. C., and Sigurdsson, K. (2017). Deleveraging Risk. *Journal of Financial and Quantitative Analysis*, 52(6):2491–2522. Publisher: Cambridge University Press.

Roll, R. (1986). The Hubris Hypothesis of Corporate Takeovers. *The Journal of Business*, 59(2):197–216. Publisher: University of Chicago Press.

Rompotis, G. G. (2005). An Empirical Comparing Investigation on Exchange Traded Funds and Index Funds Performance. SSRN Scholarly Paper ID 903110, Social Science Research Network, Rochester, NY.

Sapp, T. and Tiwari, A. (2004). Does stock return momentum explain the "smart money" effect? *Journal of Finance*, 59(6):2605–2622.

Sherrill, D. E. and Stark, J. R. (2018). ETF liquidation determinants. *Journal of Empirical Finance*, 48:357–373.

Shin, S. and Soydemir, G. (2010). Exchange-traded funds, persistence in tracking errors and information dissemination. *Journal of Multinational Financial Management*, 20(4):214–234.

Shum, P. M. and Kang, J. (2012). The Long and Short of Leveraged ETFs: The Financial Crisis and Performance Attribution. SSRN Scholarly Paper ID 1646160, Social Science Research Network, Rochester, NY.

Sirri, E. R. and Tufano, P. (1998). Costly Search and Mutual Fund Flows. *The Journal of Finance*, 53(5):1589–1622. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/0022-1082.00066.

Soggiu, M., Mateus, C., and B. Mateus, I. (2020). Do Smart Beta ETFs Deliver Persistent Performance? SSRN Scholarly Paper ID 3590995, Social Science Research Network, Rochester, NY.

Zarate, J. G., Boyadzhiev, D., and Leitao, B. (2021). Spotlight on Synthetic ETFs in Europe: A Review of Management Practices. *Morningstar Manager Research, EMEA*.