Smart Money and Mutual Fund Family

Zachary Matteucci

16th May 2022

Master's Thesis

Stockholm School of Economics

Master's Degree in Finance (MSc Finance)

Supervisor

Prof. Dr. Magnus Dahlquist

Submitted by

Zachary Matteucci

Student Nr. 42208

42208@student.hhs.se

Abstract

This paper studies the *smart money effect* within mutual fund families by looking at monthly US equity open-ended mutual fund data between 2011 and 2019. Investors' money is said to be *smart* when it is able to predict future performance. By contrast, money is said to be *dumb* when it follows a strategy that reduces the wealth of investors (when investors buy funds that perform badly and sell funds that perform well in the future). This research investigates whether mutual fund investors are *smart* and able to identify positive and negative future performance among all funds on the market *and* within fund families. Previous studies show that fund families matter for investors, who are better at picking funds within families with which they have experience. To measure the potential smart money effect, this paper first distinguishes money that chases past returns (expected to be dumb) from the money not correlated with past returns and flows (expected to be potentially smart). To do so, the fund flow is regressed on its lagged performance and flows. The fitted values of the regression are the part of the flow correlated with the fund's past performance and flows, the so-called *expected* flow, and the residuals of the regression (uncorrelated part) are the so-called *unexpected* flow. This paper finds that the *unexpected* flow is smart when it comes to all funds within the market. The unexpected flow is able to predict the market rank of a fund and whether the fund will be a star (best 5% performing fund) or a dog (worst 5%) in the following month. However, no smart money could be observed within families. Investors do not seem to be able to identify a fund's future performance relative to the other funds within the same fund family. This study also documents the dumb money effect of the expected (i.e. return-chasing) flow. Both on the market and within fund families, an increase in the *expected* flow decreases, on average, a fund's rank as well as its probability to be a star, and increases the probability of the fund to be a dog in the next period. Overall, this paper finds weak evidence of smart money within the mutual fund market, which is restricted to only the best and worst performers. However, while the results are statistically significant, the economic significance and the magnitude of the effect remain low.

Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	**Meaning**
CRSP	Center for Research in Security Prices
NAV	Net Asset Value
TNA	Total Net Assets
WRDS	Wharton Research Data Services

1 Introduction

Many authors have documented the positive relationship between mutual funds' performance and their in- and outflow (e.g. see Zheng, 2008). It has been observed that investors tend to invest (inflow) in mutual funds with good performance in the past, and sell shares (outflow) of funds with prior bad performance. This effect, the flow-performance relationship, describes the positive relationship between the mutual funds' *past* performance and the *subsequent* in- and outflows (e.g. see Roussanov et al., 2020). Presuming that investors aim to increase their wealth, it can be assumed that investors buy or sell shares in funds where they believe that the *future* performance will be positive for buy orders or negative for sell orders. Prior research has tried to answer the question of whether *current* in- and outflows are able to predict *future* performance. Are investors just investing in past winners and selling past losers (hence, chasing past performance), or are investors able to identify future performing funds?

Some researchers believe that investors are able to identify future performing funds and that investors thereby rationally and efficiently allocate their investments to funds that will perform well in the future. This effect is called the *smart money* effect (Zheng, 1999). By contrast, other studies cast doubt on the predictive power of fund flows, and argue that investors are just chasing funds with good performance in the past. Some researchers such as Frazzini and Lamont (2008) even find that investors' money can be *dumb*, hence chasing future losers and selling future winners, which is a strategy that reduces the investors' wealth. Understanding whether money is smart or not is of relevance for the following reason: the smart money effect states that investors have the ability to predict future performance. If this is true, by observing in- and outflows it would be possible to forecast future performance and build a long/short portfolio strategy. This study contributes to the work of many authors who have tried to present evidence that investors are smart and *"more rational than [...] assumed"* (Gruber, 1996, p. 783) when allocating fund investments.

The key question when looking at the smart money effect is whether investors have a superior ability to identify funds that will perform well in the future. While many research papers try to observe the effect on an individual fund level, this paper analyzes the smart money effect *within families*. The research question at hand is: do investors have the ability to identify future performing funds *within families*? In other words, within the same family, are fund in- and outflows smart and able to identify fund managers who will produce positive, respectively negative, future returns?

The reason why this paper is focusing on fund families is twofold. Firstly, the importance of fund families for the asset management industry is high, with a market particularly concentrated among a small number of big players (Ferreira and Ramos, 2009). Secondly, fund families matter for investors. Various studies show that the fund families' brand and characteristics, such as size, age, product assortment, and the presence of star funds within families, influences money flows (e.g. see Benson et al., 2008; El Ghoul and Karoui, 2020). Interestingly, Joo and Park (2011) find some evidence of smart money within families by showing that stars attract inflows for all other funds *within the same family* and tend to have persistent positive performance. Additionally, studying fund families is also relevant from an individual investor point of view. Some investors tend to stick to only one family for all their investments due to cost reasons, additional services (e.g. portfolio's risk and performance analysis) for being large clients within the same family, or good experience with one specific family (Kempf and Ruenzi, 2008a; Elton et al., 2006, 2007; Gerken et al., 2018). Gerken et al. (2018) show that investors are better at picking future performing funds within families with which they have experience. This finding could suggest that investors may be able to identify funds, within fund families, which will perform well in the future. These observations lay the ground for this research, which looks at whether smart money exists within families.

To investigate the research question at hand, this paper uses monthly observations between 2011 and 2019 of US equity open-ended mutual funds. The sample used is retrieved from the Survivor-Bias-Free US Mutual Fund database provided by the Center for Research in Security Prices (CRSP). The main variable of interest is the fund flow calculated by taking the return-adjusted *relative* change in a fund's total net assets (TNA). Thereby, the fund flow variable captures the in- and outflows of a fund as a percentage of its TNA. Based on previous literature, it seems that while not all inventors' money can be smart, some may be. To try to identify smart money, this paper therefore first identifies the return-chasing money, i.e. money that just buys past winners and sells past losers, which is assumed to be dumb money. To do so, this paper first regresses the fund flows on their lagged performance and flows (the flow regression) to identify the part of the flow correlated with the fund's past performance and flows. This correlated part is termed the *expected flow*, and is the part of the flow that can be expected in time t based on the fund's past (t-1, t-2, t-3, etc.) performance and flows. The expected flows are calculated by using the fitted values of the flow regression. By contrast, the residuals of the flow regression are the *unexpected flows*, i.e. the part of the flow uncorrelated with the fund's past performance and flows, which is assumed to be potentially smart. By splitting the fund flow into two parts, the expected and unexpected flow, this study tries to account for investors' return-chasing behavior.

This study then tests how the unexpected and expected flows are able to predict the future fund's performance relative to all other funds in the data *and* relative to the other funds within the same family. To do so, three performance variables are constructed: the rank, star, and dog variables. For each month, the rank variable ranks a fund i's performance, first compared to all other funds on the market (*market rank*) and then compared to all other funds in the same family (*family rank*). Similarly, the star and dog variables are dummy variables which equal one if a fund i belongs to the best (for *star*) or worst (for *dog*) 5% performing fund within a month, and equal zero if not. The star and dog variables are calculated once within the market (*market* star and dog, i.e. the best respectively worst 5% within the market) and once within the family (*family* star and dog, i.e. the best respectively worst 5% within the family). Finally, this study also looks at whether the expected and unexpected flows are able to, not only identify the best and worst-performing funds but also the funds in between. This is done by classifying the funds into five different groups (*quintiles*) depending on the funds' performance. The quintile regression looks at whether the expected and unexpected flow can identify: funds belonging to the best 20% (*quintile 1*), funds between the best 20% and 40% (*quintile 2*), between the best 40% and 60% (*quintile 3*), between the best 60% and 80% (*quintile 4*) and the worst 20% (*quintile 5*) performing funds, both on the whole market and within the family (*market* and *family* quintile).

This research finds weak evidence of smart money for the unexpected flow among all funds in the data set (see Table 3). The unexpected flows are able to predict the future performance of a fund relative to all other funds *within the market*. One additional unit of unexpected flow, on average, increases a fund's rank as well as its probability to be a star and decreases its chances to be a dog fund. A fund's unexpected flow, thereby, seems to be an indicator of future performance on the market. However, while the results are statistically significant (at the 0.1% level), the magnitude of the effect remains low. On average, when unexpected flows are positive: the average monthly unexpected *inflow* is 1.9% for the data sample used in this paper between 2011 and 2019. By contrast, when the unexpected flows are negative, the average monthly unexpected *outflow* equals -1.7%. Based on the results of the smart money regression in Table 3, a 1.9% respectively -1.7% unexpected flow, on average, leads to an approximate (-)0.002 change in the market rank and a (-)0.1 percentage point change in the fund's probability to be a star/dog. The effect of the average unexpected flow on the performance variables seems limited. Additionally, it is important to keep in mind that the unexpected flow is measured in the percentage of the fund's TNA. Using the average fund size in this sample, a 1.9% unexpected inflow represents 643 million US dollars and 1.7% unexpected outflow 564 million US dollars. For these reasons, the paper at hand deems the economic significance of the results to be low.

However, within fund families, the unexpected flow has a low/no predictive power for future performance (see Table 5). When regressing the lagged fund's unexpected flow on the family rank, star, and dog variables, the only statistically significant coefficient (for the family rank) even points in the opposite direction: on average, one additional unit of unexpected flow *decreases* the market rank. This effect shows signs of dumb money which chases future losers and sells future winners.

The results of the regressions also show that the dumb money effect can clearly be observed among the expected flows, both on the market and within families. One additional unit of unexpected flow decreases on average a fund's future market rank and its probability to be a star, as well as increases its chances to be a dog fund. Therefore, the expected flow, i.e. the money chasing past performance and flows, seems not to be able to identify future winners and avoid future losers. Finally, the quintile regression shows that the predictive power of smart and dumb money is restricted to the extreme performers (best and worst 20%), i.e. the big winners and the big losers (see Table 4 and Table 6).

This paper contributes to two strings of literature on mutual fund flows: (1) the smart money effect and (2) the influence of fund families on flows. Firstly, this paper shows evidence of smart money on the market using monthly data which corroborates Zheng's (1999) and Frazzini and Lamont's (2008) observations that fund flows are smart in the short term. This research shows that the unexpected flows have a predictive power for a fund's future performance relative to other funds on the market. While Zheng (1999) and Frazzini and Lamont (2008) use fund flows and excess returns, this study contributes to the research in this field by providing a more granular view on fund flows by distinguishing the expected flows from the unexpected flows. Additionally, this paper does not look at the absolute performance, as Zheng (1999) and Frazzini and Lamont (2008) do, but at the *relative* performance.

Secondly, looking at fund families, only a few researchers such as Gerken et al. (2018) try to look at the smart money effect within families. Gerken et al. (2018), when studying individual trades, show that investors have a better ability to pick funds within families with which they have experience. The current study finds contradicting evidence with Gerken et al. (2018) who show that investors make better investment for themselves when sticking to families with which they have experience. On the contrary, this research finds that this does not mean that investors are necessarily smart and able to predict future performance. Still, the contradicting results can be attributed to the differences in the data sets used. While Gerken et al. (2018) look at individual buy and sell decisions (using trade data), this study looks at the aggregated fund-level monthly in- and outflows.

This paper proceeds as follows. First, Chapter 2 provides an overview of related literature. Then, Chapter 3 describes the data set, the variables of interest, and the methodology used. Finally, Chapter 4 presents and discusses the results as well as the robustness checks and limitations of this study.

2 Overview of Related Literature

This chapter provides the theoretical background for this study. Chapter 2.1 shortly summarizes previous findings on mutual fund flows, then Chapter 2.2 introduces the smart money effect and how many empirical findings cast doubt on the investors' ability to identify good fund managers. While it does not seem probable that *all* money to mutual funds is smart, *some* money may be. Chapter 2.3 explains why this paper believes that money flows *within fund families* could potentially be smart.

2.1 Mutual Fund Flows

As a brief introduction, this chapter summarizes some empirical observations about mutual fund flows. Many studies show that mutual fund flows are strongly correlated with their past returns, an effect which is known as the *flow-performance relationship* (see Chevalier and Ellison, 1997; Sirri and Tufano, 1998; Ippolito, 2015). Roussanov et al. (2020) show that investors tend to reward funds with recent good performance with inflows and punish prior bad performance with outflows. The authors provide evidence that, in good times, investors are overly optimistic about the fund managers' ability, and excessively reward good returns with inflows. However, in bad times, investors are rather slow to adapt. This observation reflects the non-linearity or convexity of the flow-performance relationship, as observed by Chevalier and Ellison (1997): investors tend to sell past losers but not as much as they chase past winners that performed well in the past period. Zheng (2008) argues that the asymmetry between the effect of positive and negative performance on flows lies in the fact that, when performance is good, investors consider past performance as a predictor for future performance but not in times when performance is bad. Finally, another factor influencing flows to mutual funds is the visibility of a fund, hence how much marketing and advertisement is done by a fund (Gallaher et al., 2008) and how visible and mentioned a fund is in prominent media (Reuter and Zitzewitz, 2006; Kaniel and Parham, 2017).

2.2 The Smart Money Effect

Chapter 2.2.1 first presents evidence corroborating the assumption that investors are smart. Then Chapter 2.2.2 discusses the findings of authors who cast doubts on the smart money effect and consider that money is dumb, i.e. not able to identify future performance. Finally, Chapter 2.2.3 wraps up findings on whether investors' money is smart or dumb.

2.2.1 Supporting Evidence

Gruber (1996) analyzes mutual funds data between 1985 and 1994, and concludes that future performance can be partially explained by past performance. The author also finds that future performance can be predicted by observing the current fund in- and outflows resulting from past performance. Gruber (1996) argues that a group of *sophisticated* investors is able to identify future performance and by following their investments (in- and outflows), future fund performance can be predicted. Gruber (1996) distinguishes two different types of investors: the *sophisticated* or *advantaged* and the *unsophisticated* or *disadvantaged* investor. The former type of investor is more likely to identify future performance, therefore contributes more largely to the percentage of aggregated in- and outflows, and earns abnormal return. Contrarily, the disadvantaged investors are less proactive and earn lower returns due to regulations (*"institutionally disadvantaged investors"*), tax constraints, or less information (*"unsophisticated investors"*)(p. 807).

Based on Gruber (1996), Zheng (1999) formulates the smart money hypothesis. Zheng (1999) finds that investors are able to identify future short-term performance and that the aggregated fund flows have a predictive power on funds' future performance. The author argues that this effect is not a performance-chasing behavior from the side of investors. The investors are, according to Zheng (1999), smart and do not simply invest in funds that performed well in the last period but are able to identify future performing funds. Zheng (1999) observes, however, that on average funds with inflows do not beat the market with the exception of small funds. The author finds that a trading strategy which follows small funds' in- and outflows delivers superior return. This observation is also confirmed by Sawicki and Finn (2002) and Salganik (2013).

Still, Zheng (1999) identifies that a significant part of the performance can be explained by the *"repeat winner strategy"* (p. 904), which is confirmed by some studies showing that most of the effect can be explained by the momentum factor (e.g. see Bu and Lacey, 2008). As explained more in detail in Chapter 2.2.2, many authors provide evidence against the smart money effect by attributing the short-term predictive power of flows (as observed by Zheng, 1999) to the momentum effect or the persistent flow hypothesis (Jiang and Yuksel, 2017). Additionally, many researchers cast doubts on whether investors are rationally identifying good managers or if most of the time investors are just chasing past performance, especially for biased investors (see Bailey et al., 2011). However, Berk and Green (2004) state that chasing past performance is not necessarily in contradiction with a rational and competitive market. The authors argue that investors chasing performance is not an irrational effect but a sign that investors make use of the information available and consider positive performance as an evidence of a fund manager's ability. Berk and Green (2004) argue that by allocating money in a competitive market, investors invest in last period's winner until the *"expected excess returns going forward are competitive"* (p. 1271), which under a decreasing return to scale assumption explains why alpha decreases, hence is not persistent. The authors explain that with a limited supply of skilled managers, alpha is determined by the competition on the market and not by the ability of fund managers. Building on this analysis, Berk and van Binsbergen (2015) argue that fund manager skills should not be measured in terms of return (i.e. not using alpha) but in terms of value added[1]. Using this approach, the authors find that some fund managers do possess skills and are able to produce persistent returns. Additionally, Berk and van Binsbergen (2015) state that investors seem to be able to identify and reward skilled fund managers with inflows.

Wermers (2003) provides an alternative view to counter the argument that smart money can be explained by momentum. While the author confirms that mutual fund returns are highly persistent due to stock momentum and mutual fund persistent flows (see more on this in Chapter 2.2.2), this does not *completely* explain investors inflows (as suggested by Zheng, 1999). Wermers (2003) shows that while investors do react to a fund's good past performance, the effect remains longer than the fund outperforms, which the author argues to be a reputation effect on investors, showing that investors are note solely chasing last period's winners. Similar to Berk and Green (2004), Wermers (2003) finds that net returns are diluted with persistent flows to funds. While Wermers (2003) argues that money is smart when chasing past winners, the author concedes that *"the multi-year persistence in performance can be traced, in part, to the persistence inflows over multiple years thus, consumers expectations of future performance are, to some extent, self-fulfilling."* (p.37). This casts doubt on whether all flows can be explained by the smart money effect.

Finally, both Salganik (2013) and Keswani and Stolin (2008) find evidence of smart money effect even when controlling for the momentum effect using Carhart's (1997) four-factor model. More surprisingly none of them find any difference in the ability to pick funds between retail and institutional investors. This seems to be in contradiction to Gruber's (1996) hypothesis that only sophisticated investors' money is smart[2]. Salganik (2013) explains this observation with the performance persistence, which *"probably (...) represents one of the main observable attributes of the superior ability of the fund manager, while past return information is accessible and widely used by both types of investors"* (p. 23). Again, this

[1] Berk and van Binsbergen (2015) define value added as the *"the fund's gross excess return over its benchmark multiplied by assets under management"*(p. 2).

[2] Salganik (2013) argues that the classification *institutional* investors is a good proxy for *sophisticated* investors and *retail* for *unsophisticated* investors as defined by Gruber (1996).

questions the validity of the smart money effect. Following Gruber's (1996) logic, if flows indicate the ability to identify good fund managers, it can be expected that sophisticated, i.e. institutional, investors' money should be smarter than unsophisticated, i.e. retail, investors' money. If any investor is able to do so, then money may not be as smart as expected.

2.2.2 Contradicting Evidence

As introduced in Chapter 2.2.1, three main arguments cast doubt on the smart money effect and imply that money is actually not smart, hence unable to identify future performance: 1) the persistent flow effect 2) the momentum effect 3) the impact of investors' behavioral biases.

The persistent flow hypothesis, as described by Jiang and Yuksel (2017), argues that it is the in- or outflow itself that induces (short-term) future fund performance. This theory assumes that fund managers invest new money or finance redemption with their current holdings, which automatically drives prices of the fund's underlying assets up with inflows and down with outflows of capital. Coval and Stafford (2007) show that in- and outflows induce forced trading mostly within the fund's current holdings and thereby drive prices up for inflows and down for outflows. The effect is particularly strong when other funds, which hold similar positions, experience the same in- or outflows[3]. This can be costly for funds (even for inflows) and creates opportunities for outside traders[4]. Lou (2012) quantifies the effect and shows that *"fund managers sell their holdings dollar-for-dollar to meet redemptions while investing around sixty-two cents for every dollar of inflow in their existing positions"* (p. 3459).

Other authors find a similar effect when looking at flows correlated with past performance and flows. Lou (2012) decomposes fund flows into an expected and an unexpected part. The author defines the expected flow as the portion of flows that can be predicted using past flows and past performance (fitted values of the regression). The unexpected component is the difference between the expected flows and the actual flows, i.e. the residuals of the regression. Lou (2012) shows that the expected flow is able to predict the future return of the fund's underlying assets and hence the fund itself. The author concludes that *"the flow-driven return effect can fully account for mutual fund performance persistence and the smart money effect, and can partially explain stock momentum"* (Lou, 2012, p. 3457). In other words, the author concludes that the predictive power of the flow is mostly explained by the expected flow, i.e. the persistent flow effect, rather than the smart money effect. Jiang and Yuksel (2017) find similar

[3] If many funds hold, for example, a similar portfolio of bad performing stocks, then the performance for all these funds may be low, which may trigger redemption from investors. As a consequence, the fund managers may have to liquidate part of their portfolios (to give investors their money back) which will drive down the prices of the stocks even further.

[4] Coval and Stafford (2007) show that investors can earn abnormal returns by constructing a long-short portfolio of stocks that will most likely be purchased or sold by funds as a consequence of in- or outflows.

results when analyzing the unexpected (instead of the expected) component of mutual fund flow using the same methodology as Coval and Stafford (2007) and Lou (2012)[5]. The unexpected flow is the part of flows that cannot be explained by past performance and flows, therefore the unexpected flow can be used as a proxy for the part of the flow less prone to the potential persistent flow effect. Therefore, if the unexpected flow is able to predict future performance, this could be an indicator of the smart money effect (or at least counter the persistent flow argument). However, Jiang and Yuksel (2017) find that the unexpected part of flows lacks explanatory power and when controlling for the expected flows, hence looking at the unexpected flow, *"fund flows indeed lack predictive power of future fund returns"* (p. 41).

As previously mentioned, many criticisms against the smart money hypothesis argue that the effect does not hold when controlling for momentum (see e.g. Bu and Lacey, 2008). Sapp and Tiwarsi (2004) agree that momentum may be an explanation, however, they also argue that momentum itself may not be enough to refute the smart money hypothesis as investors may simply be chasing funds with high momentum exposure (and earn compensation for the exposure to the momentum risk factor). Still, the authors show that flows are chasing funds with positive past returns more intensively than funds with high momentum exposure and when looking at funds with high momentum loading only half of them experience positive inflows. This brings Sapp and Tiwarsi (2004) to the conclusion that investors are simply chasing funds' past return (not momentum exposure) and *"in doing so, they unwittingly benefit from the momentum effect in the short term"* (p. 2608). This is consistent with Song (2020), who shows that mutual fund investors do not completely account for style factor when allocating funds. Frazzini and Lamont (2008) find that by chasing performance, individual investors tend to lose money in the long run by overweighting positions in funds with high sentiment growth stocks, that will perform poorly in the future, and underweighting funds with value stocks that perform better over time. For corporate investors, Frazzini and Lamont (2008) observe that these types of investors are less prone to investor sentiment and therefore tend to do better investment decisions.

Finally, Frazzini and Lamont (2008) are one of many researchers showing the effect of behavioral biases in investors' fund picking. Using trade information from a large US broker, Bailey et al. (2011) study the effect of such biases on investors' mutual fund picking and find that biased investors[6] tend

[5] See appendix 1 for the small differences in methodology.

[6] According to Bailey et al. (2011) typical behavioral biases, that are relevant for mutual fund investors, are for example: ***the disposition effect*** *"selling winners too quickly and holding losers too long"*, ***the narrow framing bias*** *"buying and selling individual assets without considering total portfolio effects"*, ***overconfidence*** *"frequent trading plus poor performance"* and ***the local bias*** *"preference for stocks of companies geographically close to home"* (p. 2).

to make poor investment decisions and *"trend chasing appears related to behavioral biases rather than to rationally inferring managerial skills from past performance"* (p. 1). The authors find that investors chasing returns are more subject to the overconfidence bias[6] and the *"preference for gambling and speculation"* (p. 4). Additionally, the authors observe that more sophisticated individual investors are less prone to pursue a return chasing strategy[7]. Using survey data on US investors, Choi and Robertson (2020) show that among individual investors there is a profound belief that past fund performance is a *"strong evidence that its manager has good stock-picking skills"* (p. 1967) and that active mutual funds do not suffer from diseconomies of scale. Heuer et al. (2017) show that return-chasing investors tend to only focus on top performing funds[8], ignore volatility considerations (when comparing alpha for example) and thereby *"confuse risk-taking with manager skill"* (p. 605).

2.2.3 Discussion: Are Money Flows to Mutual Funds Smart or Not?

Overall, this chapter looks at both contradicting and supporting evidence for the smart money hypothesis. While both sides provide solid arguments, it can be observed that, firstly, the key literature on smart money is relatively old and new studies cast strong doubts on the existence of smart money. Secondly, while it seems difficult to give a simple answer to the question, *are money flows to mutual funds smart or not?*, one observation is that not *all* money flows can be smart. However, it also seems possible that *some* money flows could be smart. The next chapter introduces why this paper believes that money flows *within a family* could be potentially smart money.

2.3 Family Effect

This chapter presents several findings on the relationship between mutual fund flows and mutual fund families. Chapter 2.3.1 shows that belonging to a fund family has an impact on an individual mutual fund flow, in other words, families matter for a fund's in- and outflows. Finally, Chapter 2.3.2 analyzes more in detail the investors' attitude towards fund families and explains why investors may be smart within families and able to identify good fund managers within families.

2.3.1 Fund Family Characteristics and Fund Flows

Many studies show that mutual fund families matter for individual funds. When looking at Australian funds, Benson et al. (2008) observe that families' size, age and product offering matter for investors.

[7]Please note that Bailey et al. (2011) do not use the term *sophisticated investors* in exactly the same way as Gruber (1996). Bailey et al. (2011) define *sophisticated investors* as investors *"who are professionals, live near a financial center, trade options, or have well-diversified, well performing individual stock portfolios"* (p. 14).

[8]This results, according to Heuer et al. (2017), to a sample bias, which means that the investors do not realize that by looking only at the top performers the sample may not be representative for the whole population (market).

Investors' money flow seems to prefer larger and specialised fund families, while family age seems only to matter for retail, i.e. less sophisticated, investors. A common explanation, why fund families matter, is that families can share a lot of common skills, resources and information (Brown and Wu, 2016). Individual funds also benefit from their family brand and how much their family is investing in its brand through advertising. The brand effect is especially important for well-performing individual funds but also for bad performing funds where the family brand tends to shield bad funds from large outflows (Hazenberg et al., 2015)[9]. El Ghoul and Karoui (2020) find corroborating evidence for the value of the family brand while studying funds' names. The authors find that individual funds with names close to their fund family name attract more flows and the effect is stronger for large and old families, which supports the family brand effect. Taking another approach, Sialm and Tham (2015) measure the brand effect using the fund family's stock price as a proxy and find that the performance of the fund family's stock has a positive relationship with the individual fund's in- and outflow. Finally, advertisement and marketing is another good example of how families can share resources and costs. Verbeek and Huij (2007) argue that when an individual fund advertises its product, it has a direct positive effect on the overall family brand which becomes more visible to investors. Contrasting these results, Wei et al. (2011) show that the advertising spillover effect may only significantly improve fund flows for larger families (and not smaller families) and it benefits only the best-performing funds within families.

Interestingly, Brown and Wu (2016) show that while fund families may have a scale and shared resource advantage, common skills and knowledge may also result in having funds within families with overlapping positions and thereby a family *"correlation of noise in fund returns"* (p. 385). Additionally, resources within a family may not be equally distributed between individual funds. Guedj and Papastaikoudi (2003) show that families tend to allocate resources to the best-performing funds and not according to the individual funds' needs. This can be associated with the fact that the best-performing funds, i.e. star funds, are beneficial for fund families. Another way for families to favor high-performing funds is to allocate performance across individual funds. Gaspar et al. (2006) show that families *"strategically transfer performance across member funds to favor those most likely to increase overall family profits"* (p. 73), for example through the allocation of under- or overpriced initial public offerings.

Furthermore, fund families compete with each others. Massa (2003) shows that families may use non-performance-related characteristics to compete with other families or even to compensate for poor

[9] It can be noticed that this non-linear/convex effect is in line with the non-linear/convex flow-performance relationship for individual fund discussed in Chapter 2.1.

performance. The author argues that *free-switching options*[10] and increasing the fund range (offering) within families, may be an alternative way to compete on the market. Considering free-switching options, it can be argued that investors may start to look more at the funds within the same family (cheaper to switch) and compare them, which increases the competition among fund managers within the same family. This competition can also be explained by the above-mentioned asymmetrical resource allocation within families (families tend to allocate more resources to their best-performing funds or funds most likely to become star funds). Kempf and Ruenzi (2008b) show that fund managers within the same family are competing with each other in a *"family tournament"* (p. 1014) especially in large families. The authors find that fund managers adapt their portfolio risk depending on their midyear intra-family ranking. Simutin (2013) captures the family effect of having shared resources by showing that portfolio managers who deviate from *"the "average" portfolio of other funds in the same family significantly outperform managers who passively mimic their family's portfolio"* (p. 1).

Interestingly, while fund families do have an effect on individual funds, the opposite is true as well. Individual funds also have an impact on other funds within the same family (spillover effect). Nanda et al. (2004) find that star funds[11] generate inflows for themselves but also have a positive inflow effect on other funds within the same family. However, the opposite is not true, dog funds[11] do induce outflow for the individual funds but do not have, neither a statistically nor economically significant effect on other funds within the same family. Joo and Park (2011) find similar results for both stars and dogs[12] while studying the Korean fund market, and they observed that the effect can be better explained by looking at the *number* of stars in a family rather than only *if* a family has a star. The authors conclude that families with a high concentration of stars, tend to perform better in the future and have a higher chance to keep on producing stars in the future. According to Joo and Park (2011), *"this persistency in the performance of high star-fund holding ratio families supports the smart money effect in fund investment and rationalizes conventional investment practice among the general public of buying into funds based on past fund-family performance"* (p. 759). Guedj and Papastaikoudi (2003) investigate the similar effect and show that by betting on a family's stars and shorting its dogs, investors can earn a positive alpha. Similar to Joo and Park (2011), Guedj and Papastaikoudi (2003) find persistency in family stars over time and explain it by the fact that fund families reallocate resources, especially fund managers, to star funds.

[10] The option to switch between individual funds within the same family at no or reduced costs.

[11] Defined by Nanda et al. (2004) as the funds which performance for the previous 12 months belongs to the best 5%. Reciprocally, dog funds are funds which performance belongs to the worst 5%.

[12] Joo and Park (2011) use the same measure as Nanda et al. (2004) for stars and dogs, see footnote 11.

Overall, it can be observed that not only fund families matter for flows but also for performance. Families with winners tend to keep winning in the future, and the question to be answered in the next chapter is whether investors are able to identify these future winners within families.

2.3.2 Fund Family and Investor

As already mentioned in Chapter 2.2.1, some fund families try to generate incentives for investors to keep all their investments within the same fund family. As mentioned earlier, free-switching options[10] are one of them but families also sometimes provide additional services for larger clients such as portfolio analysis, access to additional information, etc. Additionally, the investors themselves may find it convenient to invest mostly in the same families and for example to obtain all their holdings information in one single account statement from one or a few fund families (Kempf and Ruenzi, 2008a). Elton et al. (2006, 2001) also mention that sticking to one family means that investors have less new information to process. Indeed, when investors are considering a new fund family, they may experience high costs when acquiring new information about a family they are not used to. On the other hand, the authors also warn that funds within the same family tend to have overlapping positions, and investors restricting all their investment to the same fund family may be exposed to correlated risk.

Another interesting observation is that investors tend to stick to families with whom they had good experiences and, similarly to fund-level observations, investors tend to quickly reward good performance with inflows but react slower with negative performance (Gerken et al., 2018). When studying trade orders, Gerken et al. (2018) show that investors are better at fund picking within families with which they have experience, than *out-of-family* fund picking. The authors argue that investors selecting *out-of-family* funds need to process a lot of new information and therefore tend to focus more on past performance (return chasing). However, when selecting *in-the-family* funds, investors are more likely to pick funds that perform better ex-post.

This resembles the familiarity bias[13], which raises the question of whether investors are purely biased and acting irrationally, or if the overweight in familiar assets is information-driven. Before discussing the conclusions of previous studies on the subject, it is important to say that most studies consider familiarity from a geographic perspective. The question at hand in this paper is the familiarity with a given fund family which is not geographically related. A geographic or fund family bias may be different at least for big fund families, who offer various and diversified products. A geographically

[13]Defined by Foad (2011) as the tendency of biased investors to invest in assets they are familiar with. Many research papers have looked at this bias from a geographic proximity perspective, hence whether investors tend to invest more in assets they are geographically close to, on a domestic (local bias) and international (home bias) level.

biased investor may be, for instance, heavily exposed to the local or domestic industry, hence to sector risk, while a fund family biased investor (in a well-diversified family) may be exposed to other risks such as the correlated risk of overlapping positions between funds in the same family. Still, it is interesting to observe that empirical studies on the geographic familiarity bias are divided: for example, Ivkovic and Weisbenner (2005) argue that investors can use local (superior) knowledge to generate abnormal returns whereas Seasholes and Zhu (2010) argue the exact opposite, and that no alpha can be expected from locally biased investment decisions. In the end, it could be argued that what matters, when trying to identify whether investors are smart or biased, is whether investors are able to persistently identify future performance or not. Therefore, while investors may be subject to the familiarity bias when picking funds within families with which they have experience, this may not exclude them to be smart.

2.3.3 Discussion: Can Investors be Smart Within a Family?

This chapter shows that individual fund flows also depend on the fund's family. Fund managers compete against other funds on the market but also against other funds within their families. Families with winners, tend to keep performing well and investors seem to be better at picking funds within families with which they have experience. While many studies cast doubts on the existence of smart money on the whole market, this study, believes that intra-family flows may have the potential to be smart and investors may be able to identify future performance within families.

3 Data and Methodology

The following chapter describes the data (Chapter 3.1), defines the key variables (Chapter 3.2), and introduces the methodology (Chapter 3.3) used in this paper.

3.1 Data

For the analysis, this paper uses the data provided by the CRSP on Survivor-Bias-Free US Mutual Fund (University of Chicago)[14] retrieved using the Wharton Research Data Services (WRDS)[15]. The CRSP database, largely used in the research on mutual funds, provides yearly, quarterly and monthly variables on traded US open-ended mutual funds.

This paper considers monthly data between 2011 and 2019[16] and focuses only on US *equity* mutual funds[17]. The US equity market is the largest and hence most liquid market according to the Investment Company Institute (2021), which makes it suitable for this study. The data is restricted to only active funds[18]. The reason is that the smart money effect lies in the (potential) ability of investors to identify funds that will perform well in the future, thus fund managers that are able to produce future positive returns. As replicating an index does not fit this definition, index funds are removed, which is consistent with other studies in the field of smart money (see e.g. Keswani and Stolin, 2008).

As suggested by Zheng (1999), some funds with different strategies may be exposed to very different risk factors. Therefore, as suggested by Jiang and Yuksel (2017) and by Elton et al. (2007) balanced, sector, and international funds are excluded[19]. Thereby, the sample narrows this study on US equity funds with a domestic focus without sector strategy to eliminate exposure to very different risks.

Based on the scope of this paper, connecting an individual fund to its family is of the highest importance for this study. The CRSP Data set offers two possible variables: the fund family name or a fund family identifier. After reviewing both variables, some inconsistencies[20] within and between the two variables are observed. Therefore, both variables are manually corrected and only the fund family

[14] See more on http://www.crsp.org/ or Elton et al. (2001) for a more in-depth description of the data.

[15] See more on https://wrds-www.wharton.upenn.edu/ and Appendix 2 for the WRDS queries.

[16] Appendix 3 shows the yearly total funds and funds per family distribution in the data set for the time period between 2000 and 2020. During the most recent years (2011-2020) it can be observed that the amount of funds per family has been stable, explaining why this research focuses on this period. Appendix 3 also explains why the data that was actually retrieved from WRDS contains more years than in the sample data used for the regressions.

[17] To select only equity funds, the Lipper Classification Code *EQ* has been used, see more on http://www.crsp.org/.

[18] To select only active funds, all index funds are excluded using the Index Fund Flag from CRSP. This paper excluded any index and index-related funds based on the CRSP classification, see more on http://www.crsp.org/.

[19] To exclude these funds, this paper used the Lipper Objective Codes, see more in Appendix 4.

[20] E.g. long family names with abbreviations are not always consistently used, for example, *JPM*, *JPMorgan*, *JPMorgan Chase*, or company names are sometimes counted twice by adding or removing *& Co.* or *Group*, etc.

identifier is used in this study.

As CRSP considers each different share class of a fund as a different observation (Wermers, 2003), this paper uses the MLINKS database, provided by Russ Wermers and the Wharton Research Data Services[21], to aggregate the observations at a fund level (as suggested by Brown and Wu, 2016)[22]. Additionally, the MFLINKS database also provides information on whether the fund has been part of a merger or not. As mergers increase or decrease a fund's TNA, funds that underwent a merger are excluded (consistent with Ammann et al., 2019).

This paper wants to look at the smart money effect within families and therefore needs to exclude small families with only a few funds. For this reason, all families that have less than ten funds in the sample are removed.[23]

While constructing the flow variable, by using the change in a fund's TNA (see Equation 2 in Chapter 3.2), this paper obtains extreme values (outliers). By having a detailed look at these values, it can be observed that they mostly occur when the fund's TNA is below ten million US dollars. This observation is consistent with Kempf and Ruenzi (2008a), who consider that observations with a TNA below 10 million are *"often unreliable"* (p. 185). Therefore all monthly observations with a TNA below ten million are removed.

To build the main variable of interest, the *unexpected flow*, this paper first runs a regression of fund i's flow on its (1-,2-,3, ..., n-)lagged flows. This requires fund data from several months and *consecutive* observations. Therefore, this study excluded funds for which less than 12-month data was available and without consecutive observations.

Finally, it is worth mentioning that in the WRDS query for the CRSP database (see Appendix 2 for detailed explanations) there are significantly more variables available for annual observations than for monthly observations. Therefore, in the first step, annual data is retrieved for all available funds within the selected time period (more than 500'000 observations with every 32 variables). Then, some of the data manipulations mentioned above are executed on the annual data to identify the relevant funds for the study at hand (exclusion of passive funds, funds with different risk factors, and funds with a merger). In a second step, the monthly data is queried for the relevant funds. This procedure enables this study to consider all funds in the CRSP population without having to handle large data files (for

[21]See more on https://wrds-www.wharton.upenn.edu/.

[22]To aggregate the data on a fund level, this study uses TNA weighted values. Hence, the fund return, for instance, is obtained by calculating the average share class return using each share class' TNA as weights.

[23]Nanda et al. (2004) exclude only single-family-funds but also have a different scope than this paper.

the same timeline, there are more than 4.4 million monthly observations).

Table 1: Summary Statistics
Table 1 gives an overview of the data set. For each year the table shows: 1) the total number of different funds 2) the total number of different families 3) the total number of TNA in billion US dollars (B) 4) the fund's average monthly return 5) the average fund's monthly flow 6) the average TNA per fund in million US dollars (M) 7) the average TNA per family in million US dollars (M).

Year	Funds	Families	TNA (B)	Avg. Month. Return (%)	Avg. Month. Flow(%)	Avg. Age (Y)	Avg. TNA per Fund (M)	Avg. TNA per Fam. (M)
2011	698	36	163,951	-0.17	0.28	14.00	23,489	455,420
2012	734	39	179,962	1.21	-0.07	14.00	24,518	461,440
2013	783	44	228,062	2.32	0.38	15.00	29,127	518,322
2014	842	47	262,535	0.68	0.28	15.00	31,180	558,586
2015	894	49	288,700	-0.06	0.19	16.00	32,293	589,184
2016	972	50	335,881	0.88	-0.11	16.00	34,556	671,761
2017	1015	53	391,244	1.46	0.08	16.00	38,546	738,197
2018	1022	51	436,546	-0.50	-0.08	17.00	42,715	855,973
2019	1014	51	452,064	1.98	-0.10	18.00	44,582	886,400

Table 1 provides summary statistics for the final data set. Overall, the sample consists of (on average between 2011 and 2019) 886 funds with an average age of 16 years and there are on average 47 fund families in the sample. The total TNA, the TNA per fund and per family always increased during the period between 2011 and 2019, which is consistent with the general development of the mutual fund market (Investment Company Institute, 2021).

3.2 Variables

This chapter presents the different variables used: the flow (Chapter 3.2.1), the expected and unexpected flow (Chapter 3.2.2), the rank (Chapter 3.2.3), the star and dog (Chapter 3.2.4), and the quintile variables (Chapter 3.2.5).

3.2.1 Flow

In order to estimate the fund's in- and outflow, this paper uses the change of a fund's TNA as a proxy (based on Zheng, 2008). The TNA is affected by the investors' in- and outflow but also by a potential fund merger (which has been accounted for in the data selection[24]) and the fund's return. When a fund has a positive (negative) return, this means the value of its assets increases (decreases) and therefore the change in TNA must be corrected for the fund's return in order to isolate fund flows.

The change in TNA, $\Delta TNA_{\mathrm{i,t}}$, is adjusted for the fund's return by multiplying the TNA at the beginning of the month *t*, hence the TNA at the end of month *t-1* (CRSP provides TNA values measured at the end of the month), $TNA_{\mathrm{i,t\text{-}1}}$, with month *t* return ($R_{\mathrm{i,t}}$). $\Delta TNA_{\mathrm{i,t}}$ is calculated using a merger-

[24]See Chapter 3.1, merged funds are excluded in the data sample using the MFLINKS database.

free database with the following equation:

$$\Delta TNA_{\text{i,t}} = TNA_{\text{i,t}} - TNA_{\text{i,t-1}}(1 + R_{\text{i,t}}) \tag{1}$$

where $TNA_{\text{i,t}}$ is the TNA in US dollars of fund i for the period t, $TNA_{\text{i,t-1}}$ is the TNA of fund i for the period $t\text{-}1$ and $R_{\text{i,t}}$ is the return of fund i for the period t. $\Delta TNA_{\text{i,t}}$ captures the absolute change of a fund's TNA. To make the comparison with funds of different sizes possible, the $Flow_{\text{i,t}}$ variable is defined as the $\Delta TNA_{\text{i,t}}$ relative to the fund's return-adjusted TNA for the previous period . Thereby, based on Berk and Green (2004), $Flow_{i,t}$ is measured using the following equation:

$$Flow_{\text{i,t}} = \frac{TNA_{\text{i,t}} - TNA_{\text{i,t-1}}(1 + R_{\text{i,t}})}{TNA_{\text{i,t-1}}(1 + R_{\text{i,t}})} = \frac{\Delta \text{TNA}_{\text{i,t}}}{TNA_{\text{i,t-1}}(1 + R_{\text{i,t}})} \tag{2}$$

where $TNA_{\text{i,t}}$ is the TNA of fund i for the period t, $TNA_{\text{i,t-1}}$ is the TNA of fund i for the period $t\text{-}1$, $R_{i,t}$ is the return of fund i for the period t and $\Delta TNA_{\text{i,t}}$ is the variable defined in Equation 1.

When observing the distribution of the flow variable, a considerably high number of outliers is observed. Many of them are linked to sudden extreme monthly changes in TNA or to outliers in the monthly fund return. Therefore, following Coval and Stafford (2007), observations with a change in TNA above 200% and below -50% (which seems appropriate given monthly observations) are removed. For the fund's return, outliers are handled by winsorizing the data at the 1% and 99% percentile.

3.2.2 Expected and Unexpected Flow

To calculate the expected and unexpected flows, this paper uses a framework similar to the one used by Lou (2012), Coval and Stafford (2007), and Jiang and Yuksel (2017)[25] with the exception that this paper does not use alpha as a performance measure. Chapter 4.1.2 explains in detail the reasons why this paper considers the results using alpha as unreliable.

For this paper, the expected and unexpected flows are defined using the fitted value (for the expected flow) and residuals (for the unexpected) of the regression using the following equation:

$$Flow_{\text{i,t}} = \sum_{k=1}^{K} b_k \times Flow_{\text{i,t-k}} + \sum_{k=1}^{K} c_k \times R_{\text{i,t-k}} + a_i + a_t + \epsilon_{\text{i,t}} \tag{3}$$

where $Flow_{\text{i,t}}$ is the flow of fund i in period t as calculated in Equation 2 in Chapter 3.2.1, b_k are

[25] See Appendix 1 for the differences in methodology between the three papers.

the coefficients of the regression of $Flow_{\mathrm{i,t}}$ on its lagged values $Flow_{\mathrm{i,t\text{-}k}}$, c_{k} are the coefficients of the regression of $Flow_{\mathrm{i,t}}$ on its lagged return $R_{\mathrm{i,t\text{-}k}}$, a_{i} is the entity fixed effect and a_{t} is the time fixed effect. $\epsilon_{\mathrm{i,t}}$ denotes the residuals of the regression. $Flow_{\mathrm{i,t}}$ is thereby regressed on its lagged values and on the fund's lagged returns. The regression has a time (year) and entity (fund) fixed effect, and the standard error is clustered at the fund level. As the goal of this regression is (only) to differentiate the part of the flow correlated with past returns and flows, no control variables are used in this regression, which is consistent with Lou (2012) and Coval and Stafford (2007).

The residuals of the regression in Equation 3, $\epsilon_{\mathrm{i,t}}$, captures the part of $Flow_{\mathrm{i,t}}$ which is neither correlated with the fund i's lagged flows nor with lagged returns. Therefore, $\epsilon_{\mathrm{i,t}}$ is here considered as the *unexpected* flow, as opposed to the *expected* flow which equals the part of $Flow_{\mathrm{i,t}}$ correlated to past flows and returns, i.e. the fitted values of the regression.

3.2.3 Rank

To measure the performance of a fund i relative to the other funds, this paper ranks the fund return for each month. This paper attributes for each rank an integer from one to the total amount of items ranked (one denotes the best-performing fund in the month). Each individual fund receives a market rank, $MarketRank_{\mathrm{i,t}}$ based on the fund i's rank in a month t compared to all other funds, and a family rank, $FamilyRank_{\mathrm{i\in f,t}}$ based on the fund i's rank in a month t compared to all other funds within the family *f*.

The $MarketRank_{\mathrm{i,t}}$ includes all the funds in the data sample for the same month t and is defined using the following equation:

$$MarketRank_{\mathrm{i,t}} = \mathrm{Rank}(Return_{\mathrm{i,t}}) \in \{1, 2, ..., \sum_{i=1}^{i} Funds_{\mathrm{i,t}}\} \tag{4}$$

where the rank function attributes to each fund i a value between one and the total amount of funds in the data sample for the same month t based on the performance of fund i relative to all other funds (one being the best-performing fund on the whole market).

The $FamilyRank_{\mathrm{i\in f,t}}$ assigns a rank based on all the other funds in the same family f for each month t and is defined using the following equation:

$$FamilyRank_{i\in f,t} = \mathrm{Rank}(Return_{i\in f,t}) \in \{1, 2, ..., \sum_{i\in f=1}^{i\in f} Funds_{i\in f,t}\} \tag{5}$$

where the rank function attributes to each fund i a value between one and the total amount of funds in the same family f for each month t based on the performance of fund i relative to all the other funds in family f (one being the best-performing fund within the family f).

To be able to compare the rank for each month and use the variables in a regression, both *MarketRank*$_{\mathrm{i,t}}$ and *FamilyRank*$_{\mathrm{i\in f,t}}$ are scaled between one and zero. For each month, the best market (family) rank is set to one, and the worst market (family) rank is set to zero, and through linear extrapolation, all remaining ranks are assigned a value between one and zero.

To scale *MarketRank*$_{\mathrm{i,t}}$ and obtain *MarketRankScaled*$_{\mathrm{i,t}}$, the following equation is used:

$$MarketRankScaled_{\mathrm{i,t}} = \frac{\max(MarketRank_{\mathrm{t}}) - MarketRank_{\mathrm{i,t}}}{\max(MarketRank_{\mathrm{t}}) - 1} \tag{6}$$

To scale *FamilyRank*$_{\mathrm{i\in f,t}}$ and obtain *FamilyRankScaled*$_{\mathrm{i\in f,t}}$, the following equation is used:

$$FamilyRankScaled_{i\in f,t} = \frac{\max(FamilyRank_{\mathrm{i\in\ f,t}}) - FamilyRank_{i\in f,t}}{\max(FamilyRank_{\mathrm{i\in\ f,t}}) - 1} \tag{7}$$

3.2.4 Star & Dog

The rank variable described in the previous chapter looks at whether smart money is able to forecast the future rank of a fund. It can be argued that it is difficult for investors to forecast the exact rank of a fund but maybe investors are able to identify the best or worst-performing fund on the market and/or in the family. Therefore, two additional variables are constructed to see whether fund flows are able to detect: 1) the best performing funds, called star funds, 2) the worst-performing funds, called dog funds. Following Nanda et al. (2004) and Joo and Park (2011), this paper defines a star fund as a fund belonging to the top 5% (best-performing fund) and reciprocally, a dog fund belongs to the bottom 5% (worst-performing fund) for a given month t[26].

Like for the rank variable, this study calculates the star variable both within the market and within the family. To calculate the star and dog variables, the scaled rank from Chapter 3.2.3 is used: funds belonging to the top 5% have a scaled rank above 0.95, and funds belonging to the bottom 5% a scaled rank below 0.05.

MarketStar$_{\mathrm{i,t}}$ is a dummy variable which equals one when fund i belongs to the top 5% funds on

[26]It exists different ways of defining whether a fund belongs to the best or worst performing funds in a given period. Another commonly used variable to identify star and dog funds is the Morningstar Rating Database (see more on www.morningstar.com or Blume, 1998). As Joo and Park (2011) find evidence of smart money within families with a star (dog) variable using a 5% threshold, this study decides to replicate the same variable for the study at hand.

the whole market for a given month t and zero if not. The variable is calculated using the following equation:

$$MarketStar_{\mathrm{i,t}} = \begin{cases} 1, & \text{if } MarketRankScaled_{\mathrm{i,t}} > 0.95 \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

Using the same logic, *MarketDog*$_{\mathrm{i,t}}$ is a dummy variable which equals one when fund i belongs to the bottom 5% funds on the whole market for a given month t and zero if not. The variable is calculated using the following equation:

$$MarketDog_{\mathrm{i,t}} = \begin{cases} 1, & \text{if } MarketRankScaled_{\mathrm{i,t}} < 0.05 \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

FamilyStar$_{\mathrm{i \in f,t}}$ is a dummy variable which equals one when fund i belongs to the top 5% funds within a family f for a given month t and zero if not. The variable is calculated using the following equation:

$$FamilyStar_{i \in f,t} = \begin{cases} 1, & \text{if } FamilyRankScaled_{i \in f,t} > 0.95 \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

Using the same logic, *FamilyDog*$_{\mathrm{i \in f,t}}$ is a dummy variable which equals one when fund i belongs to the bottom 5% funds within a family f for a given month t and zero if not. The variable calculated using the following equation:

$$FamilyDog_{i \in f,t} = \begin{cases} 1, & \text{if } FamilyRankScaled_{i \in f,t} < 0.05 \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

3.2.5 Quintile

While the star and dog variables, as described in the previous chapter, look at the extreme performance (best- and worst-performing funds), the quintile variables enable the categorization of funds in different groups depending on their performance. These variables are also used to observe whether the potential smart money effect may have a convex relationship (similar to the flow-performance relationship, see Chapter 2.1).

For each month, all funds are divided into five different groups depending on their performance. The first group (Quintile 1) groups the top 20% best-performing funds and the fifth group (Quintile 5) the worst 20% performing funds. The second, third, and fifth quintiles follow the same logic (between the best 20% and 40% best funds for Quintile 2, between the best 40% and 60% best funds for Quintile 3, between the best 60% and 80% best funds for Quintile 4). The five quintiles are calculated twice: once depending on the fund's performance relative to all other funds on the market (Market Quintiles) and once on the performance relative to the other funds within the same family f (Family Quintiles).

The variables $\mathit{MarketQuintile\ 1\ to\ 5}_{\mathrm{i,t}}$ are five different dummy variables which equal one depending on whether fund i in month t belongs to the best 20% performing funds (Quintile 1), between the best 20% and 40% best funds (Quintile 2), between the best 40% and 60% best funds (Quintile 3), between the best 60% and 80% best funds (Quintile 4) or the worst 20% performing funds on the whole market (Quintile 5), and zero if not. All five variables are defined using the following equation:

$$\mathit{MarketQuintile\ 1\ to\ 5}_{\mathrm{i,t}} = \begin{cases} 1, & \text{if } MarketRankScaled_{\mathrm{i,t}} \\ & >= 0.80 \text{ (for } \mathit{Market\ Quintile\ 1}_{\mathrm{i,t}}) \\ & < 0.80\ \&\ >= 0.60 \text{ (for } \mathit{Market\ Quintile\ 2}_{\mathrm{i,t}}) \\ & < 0.60\ \&\ >= 0.40 \text{ (for } \mathit{Market\ Quintile\ 3}_{\mathrm{i,t}}) \\ & < 0.40\ \&\ >= 0.20 \text{ (for } \mathit{Market\ Quintile\ 4}_{\mathrm{i,t}}) \\ & < 0.20 \text{ (for } \mathit{Market\ Quintile\ 5}_{\mathrm{i,t}}) \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

The variables $\mathit{FamilyQuintile\ 1\ to\ 5}_{\mathrm{i\in f,t}}$ are constructed using the same logic but using the rank within the family *f*, which are defined using the following equation:

$$
\mathit{FamilyQuintile\ 1\ to\ 5}_{i \in f,t} = \begin{cases} 1, & \text{if } FamilyRankScaled_{i \in f,t} \\ & >= 0.80 \text{ (for } \mathit{Family\ Quintile\ 1}_{\mathrm{i \in f,t}}) \\ & < 0.80 \ \& >= 0.60 \text{ (for } \mathit{Family\ Quintile\ 2}_{\mathrm{i \in f,t}}) \\ & < 0.60 \ \& >= 0.40 \text{ (for } \mathit{Family\ Quintile\ 3}_{\mathrm{i \in f,t}}) \\ & < 0.40 \ \& >= 0.20 \text{ (for } \mathit{Family\ Quintile\ 4}_{\mathrm{i \in f,t}}) \\ & < 0.20 \text{ (for } \mathit{Family\ Quintile\ 5}_{\mathrm{i \in f,t}}) \\ 0, & \text{otherwise} \end{cases} \tag{13}
$$

3.3 Methodology

As mentioned in Chapter 2.2, this paper's hypothesis is that money simply chasing past return should not be able to predict future performance, while smart money could. For this reason, the paper at hand first differentiates the *unexpected* flow from the *expected* flow based on Lou (2012), Coval and Stafford (2007) and Jiang and Yuksel (2017)[27]. The *expected* flow, thus the part of the flow correlated with past returns and past flows, is return-chasing money expected to have no predictive power over future performance. Contrarily, if smart money exists, then it is expected to be observed among the *unexpected* flow, which is not correlated with past performance and flow.

In other words, this paper expects the *unexpected* flow to be positively correlated with a fund's future performance, hence the *unexpected* flow should be:

1. positively correlated with the rank variable
   (an increase in the unexpected flow increases the fund's scaled rank)
2. positively correlated with the star variable
   (an increase in the unexpected flow increases the chance of the fund to be a star)
3. negatively correlated with the dog variable
   (an increase in the unexpected flow decreases the chance of the fund to be a dog).

The opposite observation for the *expected* flow is not necessarily expected to be true as past winners may be future winners (or the opposite). In other words, as the expected flow is correlated with past performance, if the future performance is similar to the past performance then intuitively the expected

[27] See more in Chapter 2.2.2 and Appendix 1 for the differences in methodology.

flow should be able to predict future performance. However, this is then only due to the fact that the performance is constant[28] and not necessarily because the investors are smart. However, over time the intuition is that the *expected* flow should not be able to systematically predict future performance, therefore having low or no correlation with the fund's future return. Finally, it is also interesting to look at the effect of the *fund's* flow (that means the *expected* and *unexpected* flow together) to check if the potential smart money effect of the *unexpected* flow dominates the overall fund flow and whether the smart money effect can be observed on a fund flow level.

It is worth reminding that this research looks at the potential smart money effect *within the same family*. As previously mentioned, it is expected that, while investors may or may not be able to identify future performing funds on the market, investors are expected to be able or be better at identifying good managers within a family. Therefore this paper expects the above-mentioned three hypothesized relationships between the *unexpected* flow and the three performance variable to be stronger or more significant within families than on the whole market.

Finally, it is important to mention that as the smart money effect considers whether investors are able to identify *future* performing fund manager, the lagged flow is used. That means the study looks at whether the (1) fund flow, (2) expected flow or (3) unexpected flow in the previous month (*t-1*) is able to predict the subsequent (t) rank, star, or dog variable.

This paper uses the following two equations for the market and family smart money regressions:

$$\begin{cases} MarketRankScaled_{\text{i,t}} \\ MarketStar_{\text{i,t}} \\ MarketDog_{\text{i,t}} \end{cases} = Flow_{\text{i,t-1}}^{\text{Fund or Exp. or Unexp.}} + Controls_{\text{i,t}} + a_i + a_t + \epsilon_{\text{i,t}} \tag{14}$$

Market Smart Money Regression

$$\begin{cases} FamilyRankScaled_{\text{i}\in\text{f,t}} \\ FamilyStar_{\text{i}\in\text{f,t}} \\ FamilyDog_{\text{i}\in\text{f,t}} \end{cases} = Flow_{\text{i,t-1}}^{\text{Fund or Exp. or Unexp.}} + Controls_{\text{i,t}} + a_i + a_t + a_f + \epsilon_{\text{i,t}} \tag{15}$$

Family Smart Money Regression

where ${Flow_{\text{i, t-1}}}^{\text{Fund or Exp. or Unexp.}}$ is the one-month lagged either fund, expected or unexpected flow. The $MarketRankScaled_{\text{i, t-1}}$, $MarketStar_{\text{i, t-1}}$ and $MarketDog_{\text{i, t-1}}$, respectively $FamilyRankScaled_{\text{i}\in\text{f, t-1}}$,

[28]The effect of persistent performance or Hendricks et al.'s (1993) *Hot Hands* effect is addressed in Chapter 4.5.4.

$FamilyStar_{\mathrm{i\in f,\ t-1}}$ and $FamilyDog_{\mathrm{i\in f,\ t-1}}$, are the variables described in the previous chapters. a_{i} is the fund entity fixed effect, a_{t} is the time fixed effect, and for the family regression, a_{f} is the family entity fixed effect. $\epsilon_{\mathrm{i,t}}$ denotes the residuals of the regression. All regressions control for the fund's size (using the fund's average past year TNA), the fund's age (using the amount of months between time t and the date when the fund was offered for the first time) and the fund's style (using the Lipper Objective Code, see Appendix 4). All regressions have a time (year) and entity (fund) fixed effect, and the family regression also has a family fixed effect. The standard error is always clustered at the fund level.

As a final step, this paper also performs a quintile regression using again fund, expected, and unexpected flow and the quintiles variables as described in Chapter 3.2.5. The reason is that in the main regression, this paper looks at the exact rank, the top, and bottom 5% fund's (star and dogs). It is, however, also interesting to observe the potential predictive power of different performance groups (best 20%, best 20%-40%, etc.). As explained in Chapter 2.1, many authors have observed that the flow-performance relationship is convex (e.g. Chevalier and Ellison, 1997), hence investors buy past *big* winners *a lot* and sell *large* positions of past *big* losers. It may be possible that this non-linear relationship exists for the smart money effect as well, in other words, investors may be able to better identify future *big* winners or losers than *small* winners or losers.

This paper uses the following equations to perform the quintile regressions:

$$MarketQuintile\ 1\text{-}5_{\mathrm{i,t}} = Flow_{\mathrm{i,t\text{-}1}}^{\mathrm{Fund\ or\ Exp.\ or\ Unexp.}} + a_i + a_t + Controls_{\mathrm{i,t}} + \epsilon_{\mathrm{i,t}} \tag{16}$$

Market Quintile Regression

$$FamilyQuintile\ 1\text{-}5_{\mathrm{i\in f,t}} = Flow_{\mathrm{i,t\text{-}1}}^{\mathrm{Fund\ or\ Exp.\ or\ Unexp.}} + a_i + a_t + a_f + Controls_{\mathrm{i,t}} + \epsilon_{\mathrm{i,t}} \tag{17}$$

Family Quintile Regression

where ${Flow_{\mathrm{i,t\text{-}1}}}^{\mathrm{Fund\ or\ Exp.\ or\ Unexp.}}$ is the one-month lagged either fund, expected or unexpected flow. $MarketQuintile\ 1\text{-}5_{\mathrm{i,t}}$ and $FamilyQuintile\ 1\text{-}5_{\mathrm{i\in f,t}}$ are the variables described in the previous chapters. a_{i} is the fund entity fixed effect, a_{t} is the time fixed effect, and for the family regression, a_{f} is the family entity fixed effect. $\epsilon_{\mathrm{i,t}}$ denotes the residuals of the regression. Similar to the main regression, all regressions control for the fund's size (fund's average past year TNA), the fund's age (amount of months between time t and the date when the fund was offered for the first time) and the fund's style (fund's Lipper Objective Code). Both the market and family regressions have a time (year) and entity (fund) fixed effect, and the family regression has a family fixed effect. The standard error is clustered at the fund level.

4 Results

This section summarizes the results of the different regressions described in the previous chapter. In the first step, Chapter 4.1 shows the results of the flow regression which enables this paper to define the expected and unexpected flows. Then Chapter 4.2 and 4.3 present the results of the smart money regression. Chapter 4.2 shows the results of the market-wide regression and Chapter 4.3 the results of the within-family regression. Finally, Chapter 4.4 presents the different robustness checks and Chapter 4.5 discusses the results obtained.

4.1 Flow Regression

This chapter presents the results of the flow regression. Chapter 4.1.1 presents the regressions as described in Equation 3 in Chapter 3.2.2. Chapter 4.1.2 presents the results of the flow regression using the fund's alpha instead of its returns. As explained in Chapter 4.1.2, the results from the alpha regression are not reliable enough and therefore not used for the smart money regressions.

4.1.1 Flow Regression Using Return

Based on equation 3, this paper performs several regressions of fund i's flow on its lagged flows and returns. To find the optimal amount of lagged values to use, this research performs several regressions using different amounts of lags. Table 2 shows the results of the 1-lag (1), 3-lag (2), and 6-lag (3) regressions. All regressions have a time (year) and entity (fund) fixed effect, and the standard error is clustered at the fund level. It can be observed that the coefficients of all lagged flows are always statistically significant at the 0.1% level, it is also the case for the lagged returns, except the 1-lag return which is not significant in regression (1).

Looking at Table 2, it can be observed that the current fund's in- and outflows are positively correlated with the fund's past returns and flows. This is consistent with previous research and the positive flow-performance relationship (see e.g. Chevalier and Ellison, 1997; Sirri and Tufano, 1998; Ippolito, 2015). As the flow regression aims to distinguish the expected flow (fitted value of the regression) from the unexpected flow (residuals of the regression), this paper decides to use the regression with the highest adjusted R^2 and therefore uses Regression (2) with 3-lags to define the expected and unexpected flows.

Regression (2) shows that one additional unit of return in the last period (*Return 1 Lag*), i.e. one additional percentage point (pp) lagged return leads on average to 0.012 pp increase in the fund's current flow (the flow variable is relative to the fund's TNA, see description in Chapter 3.2.1). Interestingly,

Table 2: Regression Table: Flow Regression

Table 2 shows the results of the regression described in Equation 3. Regression (1) regresses the monthly fund flow on the fund's 1-month lagged flow and return, Regression (2) regresses the fund flow on the fund's 3-month lagged flows and returns and Regression (3) regresses the fund flow on the fund's 6-month lagged flows and returns. All regressions have a time (year) and entity (fund) fixed effect, and the standard error is clustered at the fund level.

	Dependent variable:		
	Monthly Flow		
	(1)	(2)	(3)
Return 1 Lag	0.003	0.012***	0.016***
	(0.005)	(0.005)	(0.005)
Return 2 Lag		0.028***	0.032***
		(0.004)	(0.004)
Return 3 Lag		0.027***	0.035***
		(0.004)	(0.004)
Return 4 Lag			0.035***
			(0.004)
Return 5 Lag			0.024***
			(0.004)
Return 6 Lag			0.014***
			(0.003)
Flow 1 Lag	0.187***	0.131***	0.115***
	(0.015)	(0.013)	(0.014)
Flow 2 Lag		0.089***	0.068***
		(0.009)	(0.009)
Flow 3 Lag		0.071***	0.045***
		(0.007)	(0.007)
Flow 4 Lag			0.055***
			(0.007)
Flow 5 Lag			0.048***
			(0.009)
Flow 6 Lag			0.025***
			(0.007)
Entity FE	Fund	Fund	Fund
Time FE	Year	Year	Year
Stand. Error	Clust.	Clust.	Clust.
Observations	179,219	177,926	175,944
Adjusted R^2	0.116	0.117	0.112
Note:	.p<0.1;*p<0.05;**p<0.01;***p<0.001		

the magnitude of the effect is larger with the return of the 2- and 3-month lagged return (one additional unit of 2- and 3-month lagged return induces on average a 0.028 pp respectively 0.027 pp increase in the fund's inflow). Looking at the lagged flows, it can be observed that 0.131 (13.1%) of the current flow can be explained by the fund flow in the previous period, as well as 0.089 (8.9%) and 0.071 (7.1%) for the respective 2- and 3-month lagged flows. These results show how funds that experienced previous positive inflows tend to keep experiencing inflows in the future which is consistent with the persistent flow hypothesis (see Lou, 2012; Coval and Stafford, 2007).

This paper also performs a 12-month lagged regression (see appendix 5.1) which shows similar features (all coefficients are positive and statistically significant) but this does not improve the adjusted R^2 (see Table 12). This research also tries an alternative way of regressing the fund flow on its lagged flows and returns by using averages for the past year and past 4 quarters (see appendix 5.2). This produces similar results with positive and statistically significant coefficients (see Table 13). However, using averages do not increase the adjusted R^2. Therefore the 3-month lag regression is used (Regression (2) in Table 2) to differentiate the expected flow, i.e. fitted values of Regression (2), from the unexpected flow, i.e. residuals of Regression (2).

4.1.2 Flow Regression Using Alpha

As previously mentioned, this paper calcualtes the four-factor alpha based on Carhart's (1997) model as an alternative performance measure to the fund's return but finds unreliable results. This chapter briefly explains the reasons for this.

Following the identical methodology as in the previous chapter, the same flow regressions are performed using lagged fund alphas instead of returns (see the method to calculate the alphas in Appendix 6.1 and the equation for the flow regression in Appendix 6.2). In Appendix 7.1, Table 14 shows the results of the regressions of fund i's flow on its 1-month lagged (1), 3-month lagged (2) and 6-month lagged (3) flows and alphas. All regressions have a time (year) and entity (fund) fixed effect, and the standard error is clustered at the fund level. It can be observed that most coefficients are statistically significant and the adjusted R^2 is maximized using the 3-month lagged flow and alpha regression. Similar to the flow regression using returns, this study performs a 12-month lagged flow regression using alpha (see Table 15 in Appendix 7.2) as well as a flow regression using averages (see Table 16 in Appendix 7.3) but again this does not improve the adjusted R^2.

However, the results from the flow regressions using alpha (in Table 14) show an unexpected observation: some coefficients for the lagged alphas are negative and statistically significant. This contradicts the intuition that investors are chasing past returns and means that investors invest in past losers and redeem past winners. This paper proposes two different explanations for why negative coefficients are found. Before that, it is also important to mention that when looking at the fund's alphas, most of them are negative which is in line with previous research such as Elton et al. (1996).

First, an explanation could be that investors do not calculate and/or do not use alpha when measuring the performance of a fund, or they do not use alpha on a regular basis. Some authors studying fees[29] show that investors tend to be more reactive to *"in-your-face fees"* (Barber et al., 2005, p. 2095), such as front-end fees, than to operating fees or commission fees. This shows the tendency that investors are more sensitive to *"attention-grabbing information"* (Kronlund et al., 2020, p. 1) and less to information that is not directly available or needs more effort to obtain. Obviously, calculating alpha as described by Carhart (1997) using four risk factors for each month, demands resources. It is, therefore, possible that investors do not use alpha for this reason. Additionally, it can also be argued that investors use alternative ways of calculating alpha (e.g. Jensen's (1967) alpha) to measure a fund's performance, or even completely different performance measures, such as the relative performance to a specific benchmark. Finally, when looking at monthly data (which is the case in this paper), it is questionable whether investors calculate a fund's alpha on a monthly basis. It may be that investors consider alpha, just not on a monthly basis.

Secondly, it is possible that the negative coefficients result from a measurement error and/or high volatility in the data set. As previously mentioned, most alphas are negative and have a fat-tailed distribution (large skewness/kurtosis). Eventually, a larger time frame could have been used to calculate the alphas (e.g. 24 or 36 months instead of 12) to try to reduce the volatility. This, however, comes with a cost as a longer time frame is then needed for each fund. For example, taking two years of data to calculate alpha means that funds with less than three years of data are excluded (two years of lagged data are needed to calculate the alpha and then one additional year is needed to perform the lagged flow regression, see more on the timeline in Appendix 3).

For the reasons mentioned above, this paper decides not to use alpha, hence the expected and unexpected flow using alpha. Still, the results of the flow and the smart money regressions using alpha are available in Appendix 7.4 and 7.5[30]. While the flow regressions[31] as well as the smart money regressions[32] produce statistically significant results, the economic significance and interpretation remain low.

---

[29]Studies on fees seem appropriate here as it can be considered that both fees and performance are key for investors when allocating funds.

[30]See Appendix 6.3 and 6.4 for the Rank, Star and Dog variable definitions using alpha.

[31]See Appendix 7.1 for the flow regression using 1-lag, 3-lag and 6-lag values, Appendix 7.2 for the 12-lag regression and Appendix 7.3 for the flow regression using averages.

[32]See Appendix 7.4 for the market smart money regression using alpha and Appendix 7.5 for the family smart money regression using alpha.

4.2 Market Smart Money Regression

This chapter summarizes the results of the market-wide regression to first identify whether smart money can be observed on the whole market. The *whole market* is here meant as the whole data set used in this study (see Chapter 3.1), which means that the regressions in this chapter do not yet include a fund family dimension. The results from the market-wide regressions are then compared with the results from the smart money regressions within the family in Chapter 4.3. Chapter 4.2.1 first performs the regressions of the rank, star, and dog variables on the fund, expected, and unexpected flow. Then, Chapter 4.2.2 shows the results of the quintile regressions as explained in Chapter 3.3.

4.2.1 Market Main Regression

Using the fitted values (expected flow) and the residuals (unexpected flow) of the Regression (2) in Table 2 (3-lag flow regression), this paper performs nine different regressions, summarized in Table 3, based on equation 14. The regressions in Table 3 show the relationship between the 1-month lagged fund flow, expected flow, and unexpected flow, and the market rank, star, and dog variables. All regressions have a time (year) and entity (fund) fixed effect and control for fund's size (fund's average past year TNA), the fund's age (amount of months between time t and the date when the fund was offered for the first time) and the fund's style (fund's Lipper Objective Code). The standard error is clustered at the fund level. It is worth mentioning that these are not yet the regressions within families and therefore funds belonging to small families (families with less than 10 funds) are kept in the data set (see Chapter 3.1).

First, it can be observed that for all regressions the coefficients are statistically significant (at the 1% or 0.1% level) except for the Regression (8). Further, the hypotheses formulated in Chapter 3.3 are all confirmed for the unexpected flows: the unexpected flow is: a) positively correlated with the rank variable (see Regression (3)) b) positively correlated with the star variable (see Regression (6)) c) negatively correlated with the dog variable (see Regression (9)). This means that the unexpected flows seem to be able to positively predict future market ranks as well as identify future stars and avoid future dogs. These results indicate signs of the smart money effect for the unexpected flows.

When analyzing the magnitude of the results, it is important to keep in mind that all flow variables are expressed in percentage (of the fund's TNA) and that all three dependent variables, the market rank, star, and dog variables, have values between one and zero. The market rank is scaled between one and zero, which means for example that a market rank value below 0.1 means that the fund belongs to the 10% worst performing fund. The star and dog variables are dummy variables that equal one if the fund is a dog or star and zero if not.

Table 3: Regression Table: Market Smart Money Regression
Table 3 shows the results of the market-wide regression described in equation 14 using fund's return. Regression (1) regresses monthly fund's market rank on the fund's 1-month lagged flow, Regression (2) regresses monthly fund's market rank on the fund's 1-month lagged expected flow, Regression (3) regresses monthly fund's market rank on the fund's 1-month lagged unexpected flow, Regression (4) regresses monthly market star on the fund's 1-month lagged flow, Regression (5) regresses monthly market star on the fund's 1-month lagged expected flow, Regression (6) regresses monthly market star on the fund's 1-month lagged unexpected flow, Regression (7) regresses monthly market dog on the fund's 1-month lagged flow, Regression (8) regresses monthly market dog on the fund's 1-month expected lagged flow and Regression (9) regresses monthly market dog on the fund's 1-month lagged unexpected flow. All regressions have a time (year) and entity (fund) fixed effect and control for the fund's size (fund's average past year TNA), the fund's age (amount of months between time t and the date when the fund was offered for the first time) and the fund's style (fund's Lipper Objective Code). The standard error is clustered at the fund level.

	Dependent variable:								
	Market Rank			Market Star			Market Dog		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Lag Flow	0.069***			0.038***			−0.051***		
	(0.017)			(0.014)			(0.012)		
Lag Exp. Flow		−0.471***			−0.186***			0.058	
		(0.081)			(0.065)			(0.062)	
Lag Unexp. Flow			0.091***			0.047***			−0.055***
			(0.018)			(0.014)			(0.012)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Entity FE	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund
Time FE	Year	Year	Year	Year	Year	Year	Year	Year	Year
Stand. Error	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.
Observations	146,117	144,901	144,901	146,117	144,901	144,901	146,117	144,901	144,901
Adjusted R^2	0.023	0.023	0.023	0.054	0.054	0.054	0.086	0.085	0.085

Note: .p<0.1;*p<0.05;**p<0.01;***p<0.001

The results show that one additional percentage point (pp), 0.01 (1pp), of unexpected flow, on average, increases the market rank variable by 0.00091 (see Regression (3) in Table 3), increases the chance of the fund to be a star by 0.00047, i.e. 0.047pp (see Regression (6) in Table 3), and decreases the chance of the fund to be a dog by 0.00055, i.e. 0.055pp (see Regression (9) in Table 3). When looking at the average change in the unexpected flow within the data set at hand, it can be observed that when unexpected flows are positive the average monthly unexpected *inflow* is 1.9%, by contrast, when the unexpected flows are negative the average monthly unexpected *outflow* amounts -1.7%. This means that, when a fund experiences an average 1.9% unexpected inflow, its market rank is expected to increase on average by 0.0017, the chance of the fund to be a star increases by 0.1pp, and its chance to be a dog decreases by 0.1pp. On the contrary, an average unexpected outflow of 1.7% reduces the market rank by 0.0015, the star probability by 0.08pp, and increases the dog probability by 0.09pp. Overall, while the results are statistically significant the economic significance appears to be low.

Interestingly, the fund flows (sum of the expected and unexpected flow) show the same relationship as the unexpected flow (see Regressions (1) and (4)'s positive coefficient and Regression (7)'s negative coefficient), however, the magnitude of the effect is lower. In other words, the absolute value of the *fund* flow coefficients (see Regressions (1), (4), and (7)) is lower than the absolute value of the *unexpected* flow coefficients (see Regressions (3), (6), and (9)). This could indicate that the unexpected flow is better at identifying the future market rank, star, and dog than the fund flows. In other words, one additional percentage point of unexpected flow leads on average to a higher market rank, higher probability of the fund to be a star fund, and lower probability to be a dog fund, than an additional unit of fund flow.

Finally, looking at the expected flow, it can be observed that the relationship is going in the exact opposite way as the fund flow and the unexpected flow. The higher the expected flow of a fund i is, the lowest its expected average market rank and the lowest its probability to be a star (the probability to be a dog is undefined). This indicates that the expected flow, hence the return-chasing flow, seems to be dumb money that chases losers and sell winners.

Overall, the results presented in Table 3 indicate that there is an investors' ability to identify future performing funds among flows that are not chasing past returns (unexpected return). This contradicts the results of previous research (e.g. Jiang and Yuksel, 2017; Coval and Stafford, 2007), which is further discussed in Chapter 4.5.

4.2.2 Market Quintile Regression

The previous chapter shows that fund flows and especially the unexpected flows are able to identify stars and dogs as well as predict the future market rank of a fund. As mentioned in Chapter 3.3, this chapter looks at whether flows are able to identify any ranks or if, similar to the convex flow-performance relationship, flows are mostly able to predict the biggest winners and the biggest losers.

Table 4 summarizes the quintile regressions as described in equation 16. The regressions show the relationship between the 1-month lagged fund flow, expected flow, and unexpected flow, and the market quintiles 1 to 5. All regressions have a time (year) and entity (fund) fixed effect and control for the fund's size (fund's average past year TNA), the fund's age (amount of months between time t and the date when the fund was offered for the first time) and the fund's style (fund's Lipper Objective Code). The standard error is clustered at the fund level.

First of all, it is worth noticing that only the regressions using the quintiles 1 and 5, i.e. Regression (1), (2), (3), (13), (14), and (15), have statistically significant (at the 0.1% level) coefficients, with the exception of regression (5) which coefficient is significant at the 5% level. When looking only at the

Table 4: Regression Table: Market Quintile Regression

Table 4 shows the results of the market-wide regression described in equation 16. Regression (1), (4), (7), (10), and (13) regresses monthly market quintile 1 to 5 on the fund's 1-month lagged flow. Regression (2), (5), (8), (11), and (14) regresses monthly market quintile 1 to 5 on the fund's 1-month lagged expected flow. Regression (3), (6), (9), (12), and (15) regresses monthly market quintile 1 to 5 on the fund's 1-month lagged unexpected flow. All regressions have a time (year) and entity (fund) fixed effect and control for the fund's size (fund's average past year TNA), the fund's age (amount of months between time t and the date when the fund was offered for the first time) and the fund's style (fund's Lipper Objective Code). The standard error is clustered at the fund level.

	Dependent variable:														
	Quint. 1			Quint. 2			Quint. 3			Quint. 4			Quint. 5		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Lag Flow	0.076***			0.003			−0.004			−0.008			−0.066***		
	(0.023)			(0.019)			(0.021)			(0.022)			(0.021)		
Lag Exp. Flow		−0.516***			−0.175*			0.068			0.125			0.498***	
		(0.112)			(0.101)			(0.107)			(0.098)			(0.111)	
Lag Unexp. Flow			0.099***			0.011			−0.006			−0.016			−0.088***
			(0.024)			(0.019)			(0.021)			(0.022)			(0.022)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Entity FE	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund
Time FE	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year
Stand. Error	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.
Observations	146,117	144,901	144,901	146,117	144,901	144,901	146,117	144,901	144,901	146,117	144,901	144,901	146,117	144,901	144,901
Adjusted R^2	0.048	0.048	0.048	0.017	0.017	0.017	0.042	0.041	0.041	0.017	0.017	0.017	0.066	0.066	0.066

Note: .p<0.1;*p<0.05;**p<0.01;***p<0.001

regressions (1), (2), and (3), with quintile 1, as well as the regressions (13), (14), and (15), with quintile 5, this paper observes the same relationship as the one observed in the regressions using the market rank, dog and star variables (see Table 3). The fund flow and unexpected flow are positively correlated with the probability of the fund to be a future good performing fund (stars or quintile 1) and negatively correlated with the probability of the fund to be a future bad performing fund (dogs or quintile 5). On the other hand, the expected flow has the opposite relationship.

Yet, this is not surprising as the difference between the star and dog variables, and the quintile 1 and 5, is the threshold. Star and dog funds are the best respectively worst 5% performing funds and funds belonging to quintile 1 and 5 are the respectively best or worst 20% performing funds. Therefore, it can here be observed that the star and dog effect is not only valid for the 5% best or worst but also the 20% best or worst.

Regression (3) shows that one additional percentage point of unexpected flow, increases the fund's probability to be among the best 20% performing funds (quintile 1) in the next period by 0.00099, i.e. 0.099pp, and decreases the fund's probability to be among the worst 20% performing funds (quintile 5 in Regression (15)) in the next period by 0.00088, i.e. 0.088pp. Looking again at the scenario when the unexpected flow is positive, the 1.9% average monthly unexpected *inflow* increases the fund's probability to be among the best 20% performing funds by 0.0019, i.e. 0.19pp, and decreases the chance of the fund to be among the best 20% by 0.0017, i.e. 0.17pp. Looking at the scenario when the unexpected flow is negative, the 1.7% average monthly unexpected *outflow* decreases the fund's probability to be among the best 20% performing funds by 0.0017, i.e. 0.17pp, and increases the chance of the fund to be among the worst 20% by 0.0015, i.e. 0.15pp. Similar to the main regression, while the results are highly statistically significant (0.1% level) the economic significance appears to be low.

However, the most interesting element in Table 4 is that the effect of the fund and unexpected flow is only statistically significant (at the 0.1% level) for the top and bottom 20% (Quintile 1 and 5), and (almost) all other quintiles are not statistically significant (except for the coefficient in Regression (5) statistically significant at the 5% level). This seems to corroborate the intuition that the smart money effect could be a non-linear, i.e. convex, effect, similar to the flow-performance relationship as described by Chevalier and Ellison (1997). This means that investors are good at predicting future *big winners* and *big losers* but not necessarily the fund managers in between.

4.3 Family Smart Money Regression

This chapter presents the results of the smart money regression within a family. Chapter 4.3.1 first shows the family smart money regression and Chapter 4.3.2 the family quintile regression.

4.3.1 Family Main Regression

Using equation 15, this paper performs the same nine regressions as the previous chapter but this time on a family level. Table 5 shows the regressions of the 1-month lagged fund flow, expected flow and unexpected flow (resulting from the flow regression in Chapter 4.1) on the *family* rank, star, and dog variable. All regressions have a time (year) and entity (fund and family) fixed effect and control for fund's size (fund's average past year TNA), the fund's age (amount of months between time t and the date when the fund was offered for the first time) and the fund's style (fund's Lipper Objective Code). The standard error is clustered at the fund level. As these regressions look at the performance within families, funds belonging to families with less than 10 funds are removed from the data set (as explained in Chapter 3.1), which explains the drop in the total observations compared to the market-wide regressions in Chapter 4.2.

Table 5: Regression Table: Family Smart Money Regression
Table 5 shows the results of the regression within the family described in equation 15 using fund's return. Regression (1) regresses monthly fund's family rank on the fund's 1-month lagged flow, Regression (2) regresses monthly fund's family rank on the fund's 1-month lagged expected flow, Regression (3) regresses monthly fund's family rank on the fund's 1-month lagged unexpected flow, Regression (4) regresses monthly family star on the fund's 1-month lagged flow, Regression (5) regresses monthly family star on the fund's 1-month lagged expected flow, Regression (6) regresses monthly family star on the fund's 1-month lagged unexpected flow, Regression (7) regresses monthly family dog on the fund's 1-month lagged flow, Regression (8) regresses monthly family dog on the fund's 1-month expected lagged flow and Regression (9) regresses monthly family dog on the fund's 1-month lagged unexpected flow. All regressions have a time (year) and entity (fund and family) fixed effect and control for the fund's size (fund's average past year TNA), the fund's age (amount of months between time t and the date when the fund was offered for the first time) and the fund's style (fund's Lipper Objective Code). The standard error is clustered at the fund level.

	Dependent variable:								
	Family Rank			Family Star			Family Dog		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Lag Flow	-0.058^{***}			-0.028			0.028		
	(0.021)			(0.017)			(0.018)		
Lag Exp. Flow		-0.533^{***}			-0.382^{***}			0.261^{***}	
		(0.115)			(0.090)			(0.101)	
Lag Unexp. Flow			-0.036^{*}			-0.012			0.021
			(0.021)			(0.018)			(0.018)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Entity FE	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F
Time FE	Year	Year	Year	Year	Year	Year	Year	Year	Year
Stand. Error	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.
Observations	89,211	88,570	88,570	89,211	88,570	88,570	89,211	88,570	88,570
Adjusted R^2	0.027	0.027	0.027	0.055	0.056	0.056	0.087	0.087	0.087

Note: Fund & Family (F & F) .p<0.1;*p<0.05;**p<0.01;***p<0.001

It is obvious that the results are different from the market regression. The unexpected flow regressions (Regressions (3), (6), and (9)) produce only one coefficient that is statistically significant at the 5% level, while all coefficients in the expected flow regressions (Regressions (2), (5), and (8)) are all statistically significant (at the 0.1% level). For the flow regression, only Regression (1) gives a statistically significant result.

Overall, the results are inconclusive but tend to indicate the absence of any smart money effect and the existence of a dumb money effect. Most surprisingly, the statistically significant effect between the unexpected flow and the family rank (see Regression (3) in Table 5) shows that, on average, one additional percentage point of unexpected flow leads to a decrease of the market rank by 0.00036. This means that unexpected flow negatively predicts performance (dumb money). Additionally, the significant coefficients for the expected flow (Regressions (2), (5) and (8) in Table 5) all point in the wrong direction (dumb money): the flow is negatively correlated with the fund's rank and its probability to be a star, and positively correlated with the probability to be a dog. Similarly, the overall fund flow is also negatively correlated with the family rank. This means that not only is money not smart, but it is even dumb. It buys future losers and sells future winners. Chapter 4.5 discusses more in detail the implications of these results.

4.3.2 Family Quintile Regression

Similar to the market quintile regression in Chapter 4.2.2, this chapter presents the results of the same regressions using family quintiles as described in Chapter 3.2.5 and based on the Equation 17 in Chapter 3.3. Table 6 summarizes the regressions of the family quintiles 1 to 5 on the 1-month lagged fund flow, expected flow and unexpected flow. All regressions have a time (year) and entity (fund and family) fixed effect and control for fund's size (fund's average past year TNA), the fund's age (amount of months between time t and the date when the fund was offered for the first time) and the fund's style (fund's Lipper Objective Code). The standard error is clustered at the fund level.

Similar to the market quintile regressions (see Chapter 4.2.2), the only coefficients that are statistically significant are the regressions using the quintile 1 and 5 (see Regressions (1), (2), (13), (14), and (15) in Table 6) with the exception of the coefficient in the Regression (3). Concerning the unexpected flows, only the qunitile 5 regression has a statistically significant coefficient (at the 5% level) pointing in the wrong direction: an additional percentage point of unexpected flow increases the probability of the fund to be among the worst 20% performing fund (quintile 5) by 0.00051, i.e. 0.051pp (see Regression (15) in Table 17). This contradicts the hypothesis of this study and shows weak evidence of dumb money.

For both the regressions with the fund and expected flow, the coefficients indicate evidence of dumb money as well. Both flows chase the 20% worst-performing funds (quintile 5) and are negatively correlated with the probability of the fund to be among the best 20% (quintile 1) performing fund in the next month. It can also be noticed that the expected flow seems to be positively correlated with the quintile 3 variable in Regression (8). However, quintile 3 being the quintile in the middle, this observation alone may not be enough to draw any conclusion.

Moreover, it is interesting to observe the fund flow regression. The negative correlation with the quintile 1 could be expected based on the negative statistically significant relationship between the fund flow and the star variable in Table 5's Regression (1). Interestingly, while no relationship was found with the dog variable in Table 5's Regression (9), a positive and statistically significant relationship is observed between the fund flow and quintile 5. This again corroborates the absence of smart money and even the presence of dumb money.

Finally, similarly to the observation made with the market-wide quintile regression, Table 6 shows that the relationship between the fund flow and expected flow may be convex, like the flow-performance relationship. In other words, investors are especially wrong with big winners and big losers. However, while the convex relationship seems more intuitive for the smart money effect (investors are especially correct for big winners and big losers), the convexity seems more surprising when the money is actually investing in losers and selling winners.

Table 6: Regression Table: Family Quintile Regression
Table 6 shows the results of the family regression described in equation 17. Regression (1), (4), (7), (10), and (13) regresses monthly family quintile 1 to 5 on the fund's 1-month lagged flow. Regression (2), (5), (8), (11), and (14) regresses monthly family quintile 1 to 5 on the fund's 1-month lagged expected flow. Regression (3), (6), (9), (12), and (15) regresses monthly family quintile 1 to 5 on the fund's 1-month lagged unexpected flow. All regressions have a time (year) and entity (fund and family) fixed effect and control for the fund's size (fund's average past year TNA), the fund's age (amount of months between time t and the date when the fund was offered for the first time) and the fund's style (fund's Lipper Objective Code). The standard error is clustered at the fund level.

	Dependent variable:														
	Quint. 1			Quint. 2			Quint. 3			Quint. 4			Quint. 5		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Lag Flow	−0.056*			−0.016			−0.011			0.020			0.063**		
	(0.029)			(0.026)			(0.027)			(0.025)			(0.028)		
Lag Exp. Flow		−0.682***			−0.165			0.321**			0.200			0.326**	
		(0.144)			(0.143)			(0.131)			(0.146)			(0.148)	
Lag Unexp. Flow			−0.026			−0.010			−0.023			0.007			0.051*
			(0.028)			(0.026)			(0.027)			(0.025)			(0.028)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Entity FE	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F
Time FE	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year
Stand. Error	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.
Observations	89,211	88,570	88,570	89,211	88,570	88,570	89,211	88,570	88,570	89,211	88,570	88,570	89,211	88,570	88,570
Adjusted R^2	0.054	0.055	0.055	0.020	0.020	0.020	0.049	0.049	0.049	0.023	0.023	0.023	0.066	0.066	0.066

Note: Fund & Family (F & F) .p<0.1;*p<0.05;**p<0.01;***p<0.001

4.4 Robustness Checks

This chapter reports the robustness checks for the dependent variables and the sample definition. Three different checks are performed: the first one checks whether results are stable when varying the threshold for the star and dog variable (I), the second one when varying the data set and focusing on the 10% biggest families in terms of the number of funds per family (II), finally, the last one documents the results of expanding the quintile regression into a decile regression (III).

The Robustness Check I controls whether the results of the market and family smart money regression remain stable when varying the 5% threshold used for defining the star and dog variables (best, respectively worst, 5% performing fund in the month, see Chapter 3.2.4). This study, therefore, re-performs the market and family regression described in Equation 14 and 15 by using a 1% and 9% threshold for the star and dog variable[33]. Table 19 in Appendix 8.1 shows the results of the Robustness Check I for the market regression (which is compared with the main results in Table 3 in Chapter 4.2.1) and Table 20 shows the results of the family regression (which is compared with the main results in Table 5 in Chapter 4.3.1). Overall, it can be observed that for both the market and family regressions, the results are generally robust to a change in the star and dog threshold.

The Robustness Check II controls whether the family smart money effect could exist within the biggest families, i.e. the families with the highest number of funds. It can be argued that investors gather more relevant experience within a big family offering more funds, than within smaller families. Thereby, investors may be even better at the *in-the-family* fund picking (as suggested by Gerken et al., 2018) within big families than within smaller ones. For this reason, this chapter re-performs the family main regressions in Table 5 in Chapter 4.3.1 using only the biggest 10% families[34]. To do so, a new data set was created, where all families that have less than 17 funds (10% threshold in the main data set[35]) in the sample are removed. Table 21 in Appendix 8.2 reports the results of the regression based on Equation 15 using data with only the 10% biggest families. Comparing the results with the Table 5 in Chapter 4.3.1, it can be observed that the results in the family main regression are robust to the changes in the data set and no smart money effect could be observed among the 10% biggest families.

Finally, the Robustness Check III controls whether the results from the quintile regressions are robust, when using deciles instead of quintiles, and documents how convex the smart money relationship is. Using the same methodology as for calculating quintiles in Chapter 3.2.5[36], Tables 22 to 27 in Appendix 8.3 summarize the results of the decile regressions for both the market and the family. Overall, it can be observed that the results are consistent with the results of the quntile regressions (see Table 4 in Chapter 4.2.2 and Table 6 in Chapter 4.3.2). It can be noted that the convexity of the relationship is again confirmed. A good example is the market-wide regression for the unexpected flow:

[33] To do so this paper used the Equations 8, 9, 10 and 11 but changed the threshold. To find the 1% and 9% best performing funds, this paper used the scaled rank above 99% respectively above 91% threshold for stars in Equations 8 and 10. To find the 1% and 9% worst performing funds, this paper used the scaled rank below 1% respectively below 9% threshold for dogs in Equations 9 and 11.

[34] Narrowing the sample to the biggest family does not seem relevant for the market-wide smart money regression. Therefore, this robustness check is only performed for the family (and not market) smart money regression.

[35] Looking at the monthly data, on average the 10% biggest families have more than 17 funds per month within the sample described in Chapter 3.1.

[36] The market deciles 1 to 10 are calculated in a similar way as market quintiles 1 to 5, as described in Equation 12 in Chapter 3.2.5, but using 10 instead of 5 groups to classify funds based on their monthly performance. The same logic applies for family deciles 1 to 10 using Equation 13

when using quintiles (see Regressions (3) and (15) in Table 4 in Chapter 4.2.2), both quintile 1 and 5 are statistically significant at the 0.1% level and when using deciles (see Regressions (1) to (10) in Table 24) again only decile 1 and 10 are statistically significant at the 0.1% level. This example shows that the observed smart money effect is restricted to the extreme performance (biggest winners and losers).

4.5 Discussion of the Results & Limitations

This chapter summarizes and reflects on the results presented in the previous chapters, and discusses the limitations of this paper as well as potential further research topics.

4.5.1 Flow Regression

The first regression, the flow regression in Chapter 4.1, shows that fund flows are strongly correlated with the fund's past performance and past flows. This reflects the return chasing behavior of mutual fund investors and the so-called positive flow-performance relationship (see Chevalier and Ellison, 1997; Sirri and Tufano, 1998; Ippolito, 2015). Funds with good performance in the past attract new money and investors redeem their money from past losers. Additionally, the results confirm a persistency in flows: funds that benefited from positive flow in the past tend to keep experiencing inflows in the future. This can be attributed to the so-called flow persistence hypothesis (see Lou, 2012; Coval and Stafford, 2007): funds that experience inflows (outflows) invest new money in (finance redemption with) their current holdings which then drives up (down) the prices of the underlying assets which is then reflected in the future performance. This research also performs regressions with alternative performance and flows measures (e.g. averages, see Appendix 5.2) which produce similar results. Overall, the results from the flow regressions are in line with previous literature.

4.5.2 Market Smart Money Regression

Looking at the market regressions in Chapter 4.2, this paper finds some evidence of smart money. This study observes for the unexpected flow, i.e. the fund flow uncorrelated with past performance and flows, an ability to predict future performance. As summarized in Table 3, an increase in a fund's unexpected flow in the previous period (time *t-1*) by one unit, increases the rank of a fund by 0.00091, increases its chances to be a star by 0.047pp, and decreases its chance to be a dog by 0.055pp in the following period (time t). The market quintile regression in Table 4 shows that the predictive power of the unexpected flow is restricted to the 20% best- and worst-performing funds. Even further, the Robustness Check III in Appendix 8.3 shows that the effect is actually restricted to the 10% best- and worst-performing funds. As already discussed in Chapter 2.2, there is no general consensus in the literature whether smart money exists. Some authors such as Salganik (2013) and Keswani and Stolin (2008) also find evidence of smart money, while others, such as for example Coval and Stafford (2007), Lou (2012) and Sapp and Tiwarsi (2004) contradict the existence of smart money. Still, this paper needs to put the results into perspective.

Firstly, the magnitude and the significance of the effect need to be discussed. While the market smart money and the quintile regressions present strong statistically significant results (mostly around the 0.1% level), the economic significance remains low. For the unexpected regression in Table 3, the average monthly unexpected flow (1.9% for inflows and 1.7% for outflows) leads to an approximate 0.002 change in the market rank (increase for inflows, respectively decrease for outflows), 0.1pp increase for inflows (respectively decrease for outflows) in the fund's probability to be a star, and 0.1pp decrease for inflows (respectively increase for outflows) in the fund's probability to be a dog. This effect appears to be low, while it is also important to keep in mind that the unexpected flow, as all flow variables, is expressed in the percentage of the fund's TNA. With an average TNA of 33.5 billion of US dollars, one percentage point increase in the unexpected flow means on average an additional (net) 335 million US dollar inflow. For the average 1.9% unexpected *inflows*, this means on average 643 million US dollars and for the average unexpected *outflow* of 1.7%, 564 million US dollars. This paper, therefore, deems the economic significance of the results to be limited.

Secondly, both the market quintile (see Table 4) and the market decile regressions (see Chapter 4.4 and Tables 22, 23 and 24 in Appendix 8) show that investors are only smart in regards to the 20% respectively 10% best- and worst-performing funds. It is important to recall that this paper uses monthly data and all regressions are done using monthly observations. The fact that investors seem able to identify the best and worst performers raises the question of whether this is a real ability or maybe it is actually *easy* or easier for investors to identify the extreme out- and underperformer for the next month. Blake and Morey (2000) study the predictive power of Morningstar ratings[37] on future performance and find some evidence that high and low ratings may act as predictor even though the *"ratings, at best, do only slightly better than the alternative predictors in forecasting future fund performance."* (p. 451)[38]. The Morningstar ratings being relatively available[39], it may help investors to identify future big winners and losers, which does not seem to match the definition of being smart.

[37] The Morningstar rating is a 1 (worst) to 5 (best) star fund rating system provided by the company Morningstar Inc. based on a fund's past performance, risk, fees etc. See Blume (1998) for the Morningstar methodology.

[38] Blake and Morey (2000) find evidence that the lowest rankings can help to predict performance and for the highest ranking there is only *"little statistical evidence that Morningstar's highest-rated funds outperform the next-to-highest and median-rated funds"* (p. 451).

[39] Even if access to the ratings on Morningstar's website requires a subscription (see more on www.morningstar.com), Blume (1998) argues that the rankings are also regularly published on other platforms.

As a final word, it is interesting to observe that the expected flows have a predictive power but they are pointing in the wrong direction (dumb money). Higher expected flow decreases on average the rank of the fund and the probability of the fund to be a star in the next period, as well as increases the chances of the fund to be a dog. Interestingly, Frazzini and Lamont (2008) observe that individual/retail investors tend to allocate their money in funds with future low performance. The authors attribute this to the value effect and return-chasing behavior: *"money flows into mutual funds that own growth stocks [which subsequently under perform], and flows out of mutual funds that own value stocks [which subsequently over perform]"* (p. 300). As the expected flow tries to capture the return-chasing part of the flow, the results in Table 3 seem to be consistent with Frazzini and Lamont's (2008) arguments.

4.5.3 Family Smart Money Regression

The smart money regressions within the family do not produce any evidence of a smart money effect. The unexpected flow regressions in Table 5 and Table 6 have only two statistically significant coefficients, which show signs of dumb money (flows negatively correlated with future performance). It seems that the unexpected flow has no predictive power to identify future performance within the family. This contradicts this study's hypothesis. Based on Gerken et al.'s (2018) findings that investors are better at picking future performing funds within the family they have experience with, this paper's hypothesis expects that flows within families could be smart (see Chapter 2.3.2 and Chapter 2.3.3). Still, the contradiction between the results of this study and Gerken et al.'s (2018) can be explained by three main differences.

Firstly, Gerken et al. (2018) use transaction/trade level data while this study uses aggregated mutual fund data. This not only enables Gerken et al. (2018) to differentiate buy and sell orders but also to track any potential money reallocation, hence where the money comes from and where it is going. Having such data, could have, for instance, enabled this study to observe how investors are reallocating their money and target funds within families. Secondly, Gerken et al. (2018) have the ability to observe the investors' prior experience with a family, which was not possible for this study. Thirdly, Gerken et al.'s (2018) conclusion is that investors are *better* at selecting in-the-family funds than out-of-family funds, which does not necessarily mean that the investors are able to identify the *best* in-the-family funds.

Finally, the expected flow regressions show the same dumb money pattern as for the market-wide regressions (see Table 3 and Table 6) which is in line with Frazzini and Lamont's (2008) findings (see discussion in the previous chapter). There seems to be no reason to believe that the rationale of the dumb money effect is any different for the regression within the family from for the regression within the market. Nevertheless, it is interesting to observe that the dumb money effect of the return-chasing (expected) flow persists even within families.

4.5.4 Limitations and Further Research

While some of this study's limitations have already been addressed in previous chapters, two main limitations still need to be discussed: the frequency of the data set and the performance measures used.

This paper uses monthly observations for both the fund flow and performance. While Frazzini and Lamont (2008) provide strong evidence against the smart money effect, they, however, observe that within a quarter, but not beyond, the money can be smart. Still, this observation needs to be nuanced and questioned whether, in the short-term, money is actually smart. It is relevant to mention here that many research papers have observed that, in the short-term, mutual fund performance can be persistent, what Hendricks et al. (1993) call the *Hot Hands*. Even though Hendricks et al.'s (1993) theory has been challenged by studies on momentum (see Carhart, 1997), many authors have observed a performance persistency among mutual funds at least in the short-term (e.g. Vidal-García, 2013; Elton et al., 1996; Bollen and Busse, 2005). Thereby, in the short-term and given performance persistence, maybe buying the past winners and selling the past losers can be a good short-term strategy. This could potentially explain the smartness of the unexpected flow. Still, it is interesting to see that this paper documents that even in the one-month interval, return-chasing flow (the expected flow) is dumb. It is important to note that to be able to continue the interpretation, a different time horizon is needed. The limitation of this study is to only consider a one-month time span and to test this paper's hypothesis using a different time horizon could lay the ground for further research.

Lastly, the performance measures used in this paper for both the flow and the smart money regressions can be extended. For the flow regression, which aims to identify the return-chasing flows, additional performance measures used by investors, such as ratings, could be added to better capture flows chasing past performance. Concerning the smart money regressions, the rank, star, and dog variables are all relative measures based on the performance of other funds. Potential further research could try to look more into the predictive power of flows on the absolute performance or use a different benchmark. Additionally, including volatility/risk-adjusted measures or a value-added approach (as described by Berk and van Binsbergen (2015) in Chapter 2.2.1) could also expand the scope of this study.

5 Conclusion

To observe the potential smart money effect among US equity mutual funds, this paper first divides fund flows into two components: the expected flow which is the fund flow correlated with past fund performance and flows, and the unexpected flow which is not. The unexpected flow is used to observe whether investors can identify future stars (top 5%) and dogs (bottom 5%) and predict the future rank of a fund. This study also looks at whether fund flows are able to categorize funds based on their performance using quintilies (best 20%, best 20%-40%, etc.). The smart money effect is analyzed both within the whole data set, i.e. the market regression (with market rank, stars, dogs, and quintiles) and within a family, i.e. the family regression (with family rank, stars, dogs, and quintiles).

This study finds evidence of smart money within the whole market. If a fund experiences an average monthly unexpected flow (1.9% for inflows and 1.7% for outflows), this leads to an approximate 0.002 change in the market rank (increase for inflows, respectively decrease for outflows), 0.1pp increase for inflows (respectively decrease for outflows) in the fund's probability to be a star, and 0.1pp decrease for inflows (respectively increase for outflows) in the fund's probability to be a dog. The results are statistically significant (0.1% level) but lack economic magnitude and significance. Interestingly, the expected flow, hence, the flow correlated with the fund's past performance and flows (the return chasing fund flow), is dumb money: an increase in a fund's expected flow means that the fund's market rank is expected to decrease in the next period as well as its probability to be a star, and its probability to be a dog increases. Looking not only at stars and dogs, this paper also finds that both the effect of the unexpected and expected flow can only be observed for the best and worst fund quintile (20% best or worst funds). These results are robust when changing the dog and star variable thresholds as well as when using deciles (10%) instead of quintiles (20%).

On the other hand, within families, no significant smart money effect could be observed for the unexpected flow, neither when using different thresholds for the dog and star variables nor when using deciles instead of quintiles. When looking solely at the biggest fund families in terms of numbers of funds per family, still no smart money effect can be observed. When looking at whether flows are able to predict performance within families, only the dumb money effect of the expected flow (similar to the market regression) can be observed: the expected flow, on average, decreases a fund's family rank in the next month as well as its probability to be a star, and increases the fund's probability to be a dog.

Overall, this study concludes that it seems that no smart money can be observed within fund families. In the overall data set (market), evidence of smart money can be observed but only for extreme performances, hence, big winners and big losers. However, this paper must relativize its conclusion due to methodological limitations. Firstly, the monthly timeline used may be excessively short-term oriented and monthly return observations can be affected by persistent fund performance. Additionally, further performance measures could improve the split between expected and unexpected flows. Finally, this research is also limited due to the data set used. Further research could try to replicate the same methodology and test the same hypothesis using more granular data such as transaction-level data. This would enable the possibility to differentiate buy and sell orders, to observe how investors reallocate funds, and especially observe the investors' history with a given fund family.

Appendix

Appendix 1: Expected and Unexpected Flow

Expected and Unexpected Flow According to Coval and Stafford (2007, p. 483)

$$Flow_{\text{i,t}} = a + \sum_{k=1}^{K} b_k \times Flow_{\text{i,t-k}} + \sum_{h=1}^{H} c_h \times R_{\text{i,t-k}} \tag{18}$$

Expected and Unexpected Flow According to Lou (2012, p. 3469)

$$\begin{aligned} Flow_{\text{i,t}} = {} & \beta + \beta_1 \alpha_{i,t-1} + \beta_2 adjret_{i,t-1} \\ & + \beta_3 Flow_{i,t-1} + \beta_4 Flow_{i,t-2} + \beta_5 Flow_{i,t-3} + \beta_6 Flow_{i,t-4} + \epsilon_{i,t} \end{aligned} \tag{19}$$

Expected and Unexpected Flow According to Jiang and Yuksel (2017, p. 52)

$$\begin{aligned} Flow_{\text{i,t}} = {} & \alpha_t + \beta_1 Flow_{i,t-1:t-3} + \beta_2 Flow_{i,t-4:t-6} + \beta_3 Flow_{i,t-7:t-12} \\ & + \beta_4 Return_{i,t-1:t-12} + \beta_5 log(\text{TNA})_{i,t-1} + \beta_6 ExpenseRatio_{i,t-1} \\ & + \beta_7 Turnover_{i,t-1} + \beta_8 log(FamilySize)_{i,t-1} + \beta_9 Load_{i,t-1} \\ & + \beta_1 0 ReturnVol_{i,t-1} + \beta_1 1 FlowVol_{i,t-1} + \epsilon_{i,t} \end{aligned} \tag{20}$$

Appendix 2: Data and Data Query

This study uses three data sets retrieved via WRDS queries: two of them are from the CRSP data (annual and monthly observations) and one of the from the MFLINKS database.

Appendix 2.1: CRSP Annual Data

Figure 1: Appendix: WRDS Query for CRSP Annual Data
Figure 1 shows the WRDS query for the CRSP annual data.

object ▶
▼ object {24}
▶ var [29]
file : fund_summary2
qvar : crsp_fundno
beg_d : 01
beg_m : Jan
datef : YYMMDDn8.
end_d : 31
end_m : Dec
query : value
beg_yr : 1999
end_yr : 2020
▶ extra1 [1]
format : xlsx
method : 3
qvards : fund_names
address : zachary.matteucci@student.unisg.ch
datevar : DEFAULT_DATE_VARIABLE
library : crspq
compress : zip
wrdsversion : 3
file_to_upload : value
saved_query_name : MF Annual 1999-2020
query_form_page_id : 4735
query_form_page_revision : 34539

Table 7: Appendix: CRSP Annual Data Description

Table 7 gives an overview of all variables queried from the CRSP fund annual data via WRDS. Not all variables are actually used in the analysis.

Variables	Description	Value	Used
frequency	Frequency of the data set (annual, monthly, quarterly, etc)	Character	Yes - To confirm that the frequency matches the frequency in the query
fund_id	Unique identifier on a share class level	Numeric	Yes
date	Date of the observation	Date	Yes
nav	Net asset value (NAV) in million of US dollar	Numeric	No - The paper used TNA to measure the fund's asset.
nav_date	Date of the NAV measurement	Date	No - The paper used TNA to measure the fund's asset.
tna_mio	Total net asset (TNA) in million of US dollar	Numeric	Yes
tna_date	Date of the TNA measurement	Date	Yes
CUISP	CUSIP (8-digit) Identifier	Numeric	No - The fund_id is used as identifier.
portfolio_id	Identifier for a group of securities hold in one or more different fund(s)	Numeric	No - The identifier does not enable to precisely aggregate the share class level data to a fund level data.
class	Identifier to aggregate share classes on a fund level	Numeric	No - The identifier does not enable to precisely aggregate the share class level data to a fund level data.
fund_name	Name of the fund	Character	No - Not needed
ticker	Stock exchange ticker	Character	No - Not needed
mgmtcompany_name	Name of the fund management company (fund family)	Character	Yes
mgmtcompany_nbr	Identifier for the fund management company (fund family)	Character	Yes
retail	Dummy to identify retail funds	Y/N Dummy	No - Not needed
insti	Dummy to identify institutional funds	Y/N Dummy	No - Not needed
indexfund	Identifier for index fund	Character	Yes
first_offered	Date of the first time the fund was offered	Date	Yes
fee_marketing	Marketing fees measured by the 12b-1 fee	Numeric	No - Not needed
fee_mgmt	Management fees	Numeric	No - Not needed
expense_ratio	Expense ratio	Numeric	No - Not needed
tunrover_ratio	Turnover ratio	Numeric	No - Not needed
crsp_obj_cd	CRSP combination of Strategic Insights, Wiesenberger, and Lipper objective codes	Character	No - Lipper objective code used instead
strategic_code	Strategic insight objective identifier	Character	No - Lipper objective code used instead
accrual_fund_ID	Fund accrual identifier	Y/N Dummy	No - Not needed
wiesenberger_code	Wiesenberger identifier	Character	No - Lipper objective code used instead
security_type	Type of security mainly hold by the fund	Character	No - Lipper objective code used instead
lipper_class_code	Lipper classification identifier	Character	No - Lipper objective code used instead
lipper_class_name	Lipper classification code name	Character	No - Lipper objective code used instead
lipper_objective_code	Lipper objective identifier	Character	Yes
lipper_objective_name	Lipper objective code name	Character	Yes
lipper_asset_code	Lipper asset identifier	Character	No - Lipper objective code used instead

Appendix 2.2: CRSP Monthly Data

Figure 2: Appendix: WRDS Query for CRSP Monthly Data
Figure 2 shows the WRDS query for the CRSP monthly data.

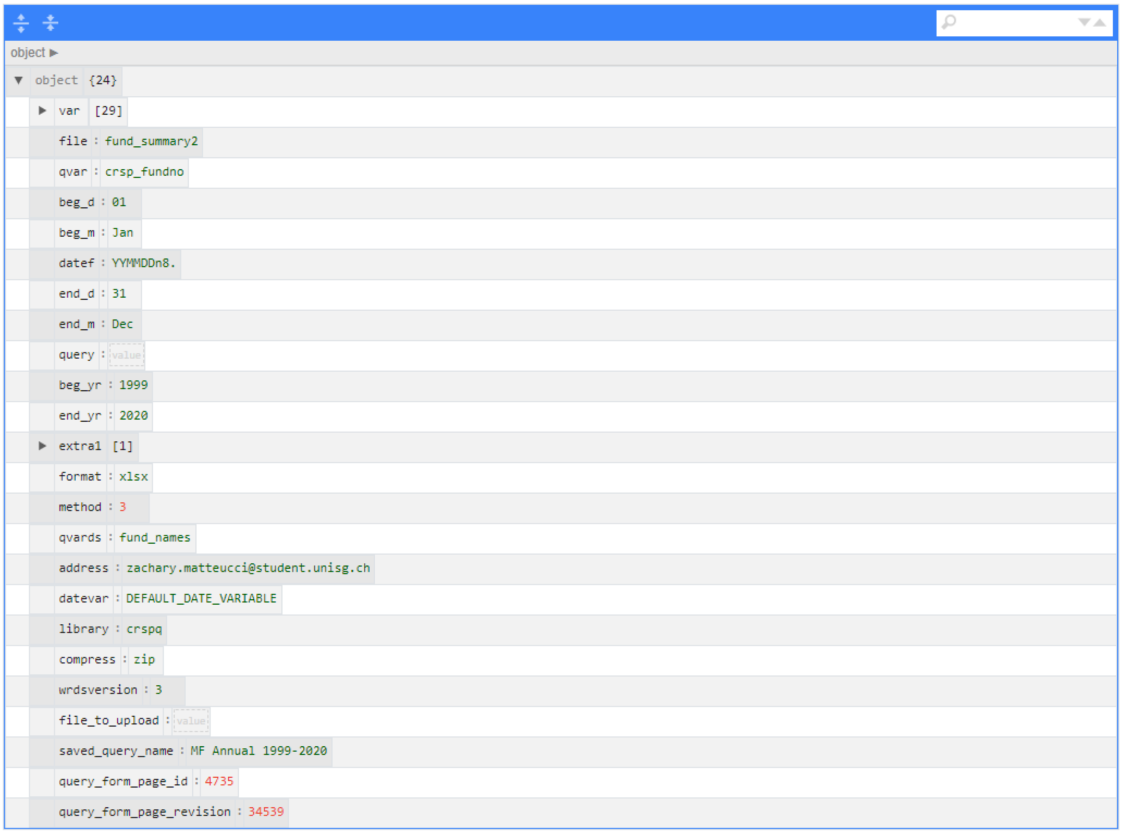


Table 8: Appendix: CRSP Monthly Data Description
Table 8 gives an overview of all variables queried from the CRSP fund monthly data via WRDS. Not all variables is actually used in the analysis.

Variables	Description	Value	Used
fund_id	Unique identifier on a share class level	Numeric	Yes
date_ymd	Date of TNA, return & NAV measurement	Date	Yes
tna_monthend	Total net asset in million of US dollar	Numeric	Yes
return_monthend	Fund's return	Numeric	Yes
nav	Net asset value in millions of dollar	Numeric	No - The paper used TNA to measure the fund's asset.

Appendix 2.3: MFLINKS Data

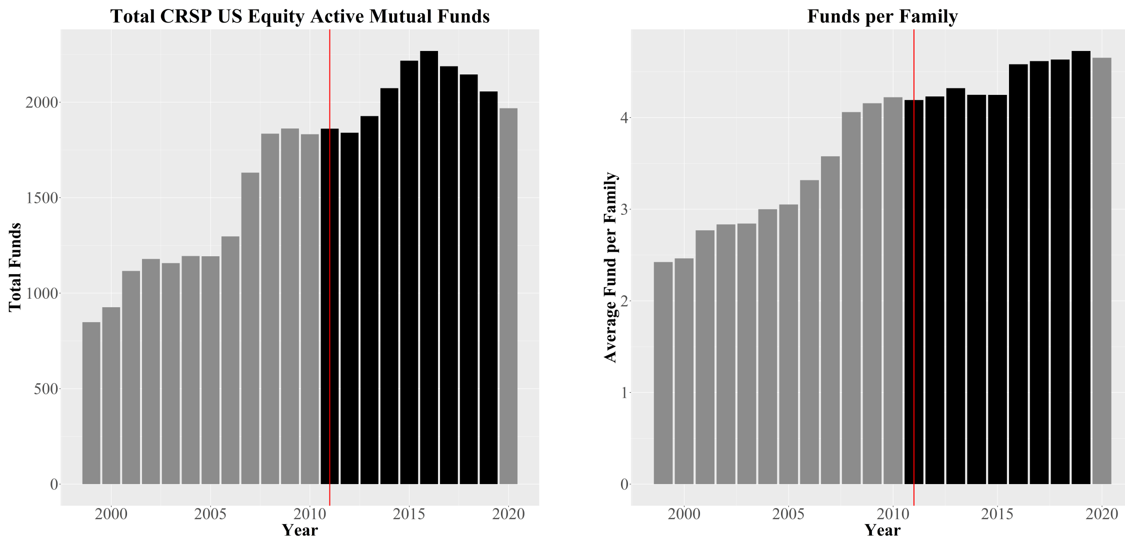
Figure 3: Appendix: WRDS Query for MFLINKS Data
Figure 3 shows the WRDS query for the MFLINKS data.

Table 9: Appendix: MFLINKS Data Description
Table 9 gives an overview of all variables queried from the MFLINKS data via WRDS. Not all variables is actually used in the analysis.

Variables	Description	Value	Used
fund_id	Unique identifier on a share class level	Numeric	Yes
wrd_nr	Unique identifier on a fund level	Numeric	Yes
mflink_cuisp	CUSIP (8-digit) identifier	Numeric	No - Not needed
mflink_fund_name	Name of the fund	Character	No - Not needed
ticker	Stock exchange ticker	Character	No - Not needed
mflink_cuisp_fund	CUSIP (9-digit) Identifier	Numeric	No - Not needed
merger	In case of merger, the variable maps the other fund(s) involved.	Numeric	Yes

Appendix 3: Timeline

Figure 4: Appendix: CRSP Annual Data Overview per Year
Both figure 4a and 4b show fund-level data of US equity-only active mutual funds excluding funds that went through a merger between 1999 and 2020 (as described in Chapter 3.1). Figure 4a shows how many different funds are available in the data set per year and 4b plots the average funds per mutual fund family per year. It can be observed that during the timeline selected the average amount of funds per family remained relatively stable.



(a) Total CRSP US Equity Open-Ended Mutual Funds **(b)** Funds per Family

Table 10: Appendix: Timeline Overview
Table 10 describes the timeline of the study at hand. As explained in Chapter 3.2, many variables used for the study are lagged variables. Therefore, to create the data sample covering 2011 until 2019, this paper needs data at least starting from 2010. However, to calculate the 12-month lagged fund's alpha, this paper needs one additional year, hence data starting from 2009. As the alpha for one month is calculated using the 12 previous monthly return (see Appendix 6.1). As the study needs 1-year lagged alpha, two additional years are needed to calculate the data set. This explains why to study the time period between 2011 and 2019, data starting from 2009 is retrieved. The year 2008 and 2020 are both excluded due to special economic situation on the financial marker: the 2008 Financial Crisis and 2020 COVID-19 (SARS-CoV-2) Crisis.

Timeline Overview				
2008	**2009**	**2010**	**2011-2019**	**2020**
Excluded due to the 2008 Financial Crisis	Included to measure alpha	Included to get lagged alpha	Sample used	Excluded due to the COVID-19 Crisis

Appendix 4: Excluding Funds by Style

Table 11: Appendix: Fund Styles

Table 11 shows which objective code are removed from, respectively kept in, the data sample. The Lipper objective and classification code describe the investment strategy of a fund based on an analysis of the fund's prospectus(CRSP, 2018). The objective code categorizes a fund's investments into different strategies (e.g. Global Funds, Growth Funds, etc) and the classification code breaks down the objective code into subcategories (CRSP, 2018). After reviewing the different objective and classification code, this paper decides to exclude funds using the Lipper objective code as there is no additional value to use a more granular data (hence no value in using the classification code). To select the relevant objectiv codes, this paper uses the overview in the *Appendix A: Data Code Listing* of CRSP (2018) and keeps only domestic equity funds without sector focus. CRSP offers alternative style code such as the Wiesenberger Objective codes and the Strategic Insight Objective codes. However, these two are not fully available for this paper's relevant timeline (CRSP, 2018).

Fund Styles in the Sample

Lipper Objective Code	Lipper Objective Name
ABR	Absolute Return Funds
CA	Capital Appreciation Funds
DL	Diversified Leverage Funds
DL	Equity Leverage Funds
DSB	Dedicated Short Bias Funds
EI	Equity Income Funds
EMN	Alternative Equity Market Neutral Funds
EMN	Equity Market Neutral Funds
G	Growth Funds
GI	Growth & Income Funds
GI	Growth And Income Funds
LSE	Alternative Long/Short Equity Funds
LSE	Long/Short Equity Funds
MC	Mid-Cap Funds
MR	Micro-Cap Funds
SG	Small-Cap Funds
SP	S&P 500 Index Objective Funds

Funds Styles Removed from the Sample

Lipper Objective Code	Lipper Objective Name
AE	Alternative Energy Funds
AED	Alternative Event Driven Funds
AGM	Alternative Global Macro Funds
AMS	Alternative Multi-Strategy Funds
AU	Gold Oriented Funds
AU	Precious Metals Equity Funds
AU	Precious Metals Funds
B	Balanced Funds
BM	Basic Materials Funds
BT	Balanced Target Maturity Funds
CG	Consumer Goods Funds
CH	China Region Funds
CMD	Commodities Funds
CS	Consumer Services Funds
CV	Convertible Securities Funds
EM	Emerging Markets Funds
EMM	Emerging Markets Mixed-Asset Funds
EMP	Energy Mlp Funds
EU	European Region Funds
FM	Frontier Markets Funds
FS	Financial Services Funds
FX	Flexible Portfolio Funds
GFS	Global Financial Services Funds
GH	Global Health/Biotechnology Funds
GIF	Global Infrastructure Funds
GL	Global Funds
GNR	Global Natural Resources Funds
GRE	Global Real Estate Funds
GS	Global Small-Cap Funds
GTK	Global Science/Technology Funds
GX	Global Flexible Port Funds
H	Health/Biotechnology Funds
I	Income Funds
ID	Industrials Funds
IF	International Funds
INR	India Region Funds
IRE	International Real Estate Funds
IS	International Small-Cap Funds
JA	Japanese Funds
LT	Latin American Funds
MFF	Alternative Managed Futures Funds
MFF	Managed Futures Funds
NR	Natural Resources Funds
OS	Options Arbitrage/Opt Strategies Funds
PC	Pacific Region Funds
RE	Real Estate Funds
RR	Real Return Funds

Appendix 5: Flow Regression

Appendix 5.1: Flow Regression (12 Lag)

Table 12: Appendix: Flow Regression 12-Lag

Table 12 shows the results of the regression described in equation 3. Regression (1) regresses the monthly fund flow on the fund's 1-month lagged flow and return, Regression (2) regresses the fund flow on the fund's 3-month lagged flows and returns, Regression (3) regresses the fund flow on the fund's 6-month lagged flows and returns and Regression (4) regresses the fund flow on the fund's 12-month lagged flows and returns. All regressions have a time (year) and entity (fund) fixed effect, and the standard error is clustered at the fund level.

	Dependent variable:			
	Monthly Flow			
	(1)	(2)	(3)	(4)
Return 1 Lag	0.003	0.012***	0.016***	0.025***
Return 2 Lag		0.028***	0.032***	0.039***
Return 3 Lag		0.027***	0.035***	0.044***
Return 4 Lag			0.035***	0.047***
Return 5 Lag			0.024***	0.040***
Return 6 Lag			0.014***	0.034***
Return 7 Lag				0.033***
Return 8 Lag				0.025***
Return 9 Lag				0.022***
Return 10 Lag				0.024***
Return 11 Lag				0.019***
Return 12 Lag				0.033***
Flow 1 Lag	0.187***	0.131***	0.115***	0.107***
Flow 2 Lag		0.089***	0.068***	0.057***
Flow 3 Lag		0.071***	0.045***	0.038***
Flow 4 Lag			0.055***	0.041***
Flow 5 Lag			0.048***	0.037***
Flow 6 Lag			0.025***	0.012
Flow 7 Lag				0.022***
Flow 8 Lag				0.006
Flow 9 Lag				0.011**
Flow 10 Lag				0.019***
Flow 11 Lag				0.007
Flow 12 Lag				0.025***
Entity FE	Fund	Fund	Fund	Fund
Time FE	Year	Year	Year	Year
Stand. Error	Clust.	Clust.	Clust.	Clust.
Observations	179,219	177,926	175,944	171,856
Adjusted R^2	0.116	0.117	0.112	0.102
Note:		.p<0.1;*p<0.05;**p<0.01;***p<0.001		

Appendix 5.2: Flow Regression Average

Table 13: Appendix: Flow Regression Average
Table 13 shows the results of the flow regression using averages. Regression (1) regresses the monthly fund flow on the fund's 1-year average lagged flow and return. Regression (2) regresses the fund flow on the fund's 4 lagged quarter average flows and returns. All regressions have a time (year) and entity (fund) fixed effect, and the standard error is clustered at the fund level.

	Dependent variable:	
	Monthly Flow	
	(1)	(2)
Average Return Lag 1 Year	0.466***	
	(0.032)	
Average Flow Lag 1 Year	0.349***	
	(0.014)	
Average Return Lag 1 Quarter		0.108***
		(0.010)
Average Return Lag 2 Quarter		0.123***
		(0.009)
Average Return Lag 3 Quarter		0.082***
		(0.009)
Average Return Lag 4 Quarter		0.078***
		(0.010)
Average Flow Lag 1 Quarter		0.203***
		(0.018)
Average Flow Lag 2 Quarter		0.088***
		(0.011)
Average Flow Lag 3 Quarter		0.037***
		(0.010)
Average Flow Lag 4 Quarter		0.050***
		(0.008)
Entity FE	Fund	Fund
Time FE	Year	Year
Stand. Error	Clust.	Clust.
Observations	171,856	171,856
Adjusted R^2	0.092	0.099
Note:	.p<0.1;*p<0.05;**p<0.01;***p<0.001	

Appendix 6: Alpha Variables

Appendix 6.1: Alpha Definition

To measure the fund's alpha, this paper uses the four-factor model as described by Carhart (1997) using the following formula:

$$R_{\mathrm{i,t}} - Rf_{\mathrm{t}} = \alpha_{i,t} + \beta_{i,t}^{\mathrm{MKT}} MKT_t + \beta_{i,t}^{\mathrm{SMB}} SMB_t + \beta_{i,t}^{\mathrm{HML}} HML_t + \beta_{i,t}^{\mathrm{UMD}} UMD_t \tag{21}$$

where $\mathrm{R_{i,t}}$ is the monthly return of fund i for the period t and $\mathrm{Rf_t}$ is the risk-free rate for the period t. $\mathrm{MKT_t}$, $\mathrm{SMB_t}$, $\mathrm{HML_t}$ and $\mathrm{UMD_t}$ are respectively the portfolios mimicking: the market (MKT), the size risk factor (Small Minus Down, SMB), the book-to-market risk factor (High Minus Low, HML) and the momentum risk factor portfolio (Up Minus Down, UMD)[40]. To calculate the fund's alpha, this paper uses the previous 12 months.

Appendix 6.2: Expected and Unexpected Flows using Alpha

$$Flow\_Alpha_{\mathrm{i,t}} = \sum_{k=1}^{K} b_k \times Flow_{\mathrm{i,t\text{-}k}} + \sum_{k=1}^{K} c_h \times \alpha_{\mathrm{i,t\text{-}k}} + \epsilon_{\mathrm{i,t}} \tag{22}$$

Appendix 6.3: Rank Variable using Alpha

$$MarketRank\_Alpha_{\mathrm{i,t}} = \mathrm{Rank}(\alpha_{\mathrm{i,t}}) \in \{1, 2, ..., \sum Funds_{\mathrm{i,t}}\} \tag{23}$$

$$FamilyRank\_Alpha_{\mathrm{i\in f,t}} = \mathrm{Rank}(\alpha_{\mathrm{i\in f,t}}) \in \{1, 2, ..., \sum Funds_{\mathrm{i\in f,t}}\} \tag{24}$$

$$MarketRankScaled\_Alpha_{\mathrm{i,t}} = \frac{\max(MarketRank\_Alpha_{\mathrm{t}}) - MarketRank\_Alpha_{\mathrm{i,t}}}{\max(MarketRank\_Alpha_{\mathrm{t}}) - 1} \tag{25}$$

$$FamilyRankScaled\_Alpha_{\mathrm{i\in f,t}} = \frac{\max(FamilyRank\_Alpha_{\mathrm{i\in f,t}}) - FamilyRank\_Alpha_{\mathrm{i\in f,t}}}{\max(FamilyRank\_Alpha_{\mathrm{i\in f,t}}) - 1} \tag{26}$$

[40] For more information on the different risk factors see Fama and French (1992) and Carhart (1997). All data for the risk factors and the risk-free rates are retrieved from Prof. Dr. French website https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Appendix 6.4: Star & Dog Variables using Alpha

$$MarketStar\_Alpha_{\text{i,t}} = \begin{cases} 1, & \text{if } MarketRankScaled\_Alpha_{\text{i,t}} > 0.95 \\ 0, & \text{otherwise} \end{cases} \tag{27}$$

$$MarketDog\_Alpha_{\text{i,t}} = \begin{cases} 1, & \text{if } MarketRankScaled\_Alpha_{\text{i,t}} < 0.05 \\ 0, & \text{otherwise} \end{cases} \tag{28}$$

$$FamilyStar\_Alpha_{\text{i}\in\text{f,t}} = \begin{cases} 1, & \text{if } FamilyRankScaled\_Alpha_{\text{i}\in\text{f,t}} > 0.95 \\ 0, & \text{otherwise} \end{cases} \tag{29}$$

$$FamilyDog\_Alpha_{\text{i}\in\text{f,t}} = \begin{cases} 1, & \text{if } FamilyRankScaled\_Alpha_{\text{i}\in\text{f,t}} < 0.05 \\ 0, & \text{otherwise} \end{cases} \tag{30}$$

Appendix 6.5: Smart Money Regression using Alpha

$$\begin{cases} MarketRankScaled\_Alpha_{\text{i,t}} \\ MarketStar\_Alpha_{\text{i,t}} \\ MarketDog\_Alpha_{\text{i,t}} \end{cases} = Flow\_Alpha_{\text{i,t-1}}^{\text{Fund or Exp. or Unexp.}} + Controls_{\text{i,t}} + a_i + a_t + \epsilon_{\text{i,t}} \tag{31}$$

Market Smart Money Regression using Alpha

$$\begin{cases} FamilyRankScaled\_Alpha_{\text{i}\in\text{f,t}} \\ FamilyStar\_Alpha_{\text{i}\in\text{f,t}} \\ FamilyDog\_Alpha_{\text{i}\in\text{f,t}} \end{cases} = Flow\_Alpha_{\text{i,t-1}}^{\text{Fund or Exp. or Unexp.}} + Controls_{\text{i,t}} + a_i + a_t + a_f + \epsilon_{\text{i,t}} \tag{32}$$

Family Smart Money Regression using Alpha

Appendix 7: Alpha Regressions

Appendix 7.1: Flow Regression using Alpha

Table 14: Appendix: Flow Regression using Alpha
Table 14 shows the results of the regression described in equation 22. Regression (1) regresses monthly the fund flow on the fund's 1-month lagged flows and alphas, Regression (2) regresses the fund flow on the fund's 3-month lagged flows and alpha and Regression (3) regresses the fund flow on the fund's 6-month lagged flows and alphas. All regressions have a time (year) and entity (fund) fixed effect, and the standard error is clustered at the fund level.

	Dependent variable:		
	Monthly Flow		
	(1)	(2)	(3)
Alpha 1 Lag	0.243***	0.174***	0.157***
	(0.012)	(0.020)	(0.020)
Alpha 2 Lag		−0.048**	−0.066***
		(0.023)	(0.024)
Alpha 3 Lag		0.090***	0.061**
		(0.022)	(0.026)
Alpha 4 Lag			0.0003
			(0.022)
Alpha 5 Lag			0.059***
			(0.018)
Alpha 6 Lag			−0.012
			(0.016)
Flow 1 Lag	0.185***	0.130***	0.116***
	(0.015)	(0.013)	(0.014)
Flow 2 Lag		0.088***	0.068***
		(0.009)	(0.009)
Flow 3 Lag		0.069***	0.045***
		(0.007)	(0.007)
Flow 4 Lag			0.055***
			(0.007)
Flow 5 Lag			0.046***
			(0.009)
Flow 6 Lag			0.023***
			(0.007)
Entity FE	Fund	Fund	Fund
Time FE	Year	Year	Year
Stand. Error	Clust.	Clust.	Clust.
Observations	179,219	177,926	175,944
Adjusted R^2	0.118	0.119	0.113
Note:	.p<0.1;*p<0.05;**p<0.01;***p<0.001		

Appendix 7.2: Flow Regression 12-Lag using Alpha

Table 15: Appendix: Flow Regression 12-Lag using Alpha
Table 15 shows the results of the regression described in equation 22. Regression (1) regresses the monthly fund flow on the fund's 1-month lagged flows and alphas, Regression (2) regresses the fund flow on the fund's 3-month lagged flows and alphas, Regression (3) regresses the fund flow on the fund's 6-month lagged flows and alphas and Regression (4) regresses the fund flow on the fund's 12-month lagged flows and alphas. All regressions have a time (year) and entity (fund) fixed effect, and the standard error is clustered at the fund level.

	Dependent variable:			
	Monthly Flow			
	(1)	(2)	(3)	(4)
Alpha 1 Lag	0.243***	0.174***	0.157***	0.155***
Alpha 2 Lag		−0.048**	−0.066***	−0.045*
Alpha 3 Lag		0.090***	0.061**	0.046*
Alpha 4 Lag			0.0003	−0.016
Alpha 5 Lag			0.059***	0.045**
Alpha 6 Lag			−0.012	−0.034*
Alpha 7 Lag				−0.011
Alpha 8 Lag				−0.047**
Alpha 9 Lag				0.056**
Alpha 10 Lag				−0.017
Alpha 11 Lag				0.219***
Alpha 12 Lag				−0.166***
Flow 1 Lag	0.185***	0.130***	0.116***	0.109***
Flow 2 Lag		0.088***	0.068***	0.059***
Flow 3 Lag		0.069***	0.045***	0.040***
Flow 4 Lag			0.055***	0.042***
Flow 5 Lag			0.046***	0.038***
Flow 6 Lag			0.023***	0.012
Flow 7 Lag				0.022***
Flow 8 Lag				0.005
Flow 9 Lag				0.010**
Flow 10 Lag				0.017***
Flow 11 Lag				0.006
Flow 12 Lag				0.023***
Entity FE	Fund	Fund	Fund	Fund
Time FE	Year	Year	Year	Year
Stand. Error	Clust.	Clust.	Clust.	Clust.
Observations	179,219	177,926	175,944	171,856
Adjusted R^2	0.118	0.119	0.113	0.102
Note:	.p<0.1;*p<0.05;**p<0.01;***p<0.001			

Appendix 7.3: Flow Regression Average using Alpha

Table 16: Appendix: Flow Regression Average using Alpha
Table 16 shows the results of the flow regression using averages. Regression (1) regresses the monthly fund flow on the fund's 1-year average lagged flow and alpha. Regression (2) regresses the fund flow on the fund's 4 lagged quarter average flows and alphas. All regressions have a time (year) and entity (fund) fixed effect, and the standard error is clustered at the fund level.

	Dependent variable:	
	Monthly Flow	
	(1)	(2)
Average Alpha Lag 1 Year	0.188***	
	(0.013)	
Average Flow Lag 1 Year	0.348***	
	(0.014)	
Average Alpha Lag 1 Quarter		0.169***
		(0.025)
Average Alpha Lag 2 Quarter		−0.030
		(0.029)
Average Alpha Lag 3 Quarter		−0.005
		(0.028)
Average Alpha Lag 4 Quarter		0.077***
		(0.027)
Average Flow Lag 1 Quarter		0.208***
		(0.018)
Average Flow Lag 2 Quarter		0.090***
		(0.011)
Average Flow Lag 3 Quarter		0.034***
		(0.010)
Average Flow Lag 4 Quarter		0.045***
		(0.008)
Entity FE	Fund	Fund
Time FE	Year	Year
Stand. Error	Clust.	Clust.
Observations	171,856	171,856
Adjusted R^2	0.091	0.099
Note:	.p<0.1;*p<0.05;**p<0.01;***p<0.001	

Appendix 7.4: Market Smart Money Regression using Alpha

Table 17: Appendix: Market Smart Money Regression using Alpha
Table 17 shows the results of the market-wide regression described in equation 31 using fund's alpha. Regression (1) regresses monthly fund's market rank on the fund's 1-month lagged flow, Regression (2) regresses monthly fund's market rank on the fund's 1-month lagged expected flow, Regression (3) regresses monthly fund's market rank on the fund's 1-month lagged unexpected flow, Regression (4) regresses monthly market star on the fund's 1-month lagged flow, Regression (5) regresses monthly market star on the fund's 1-month lagged expected flow, Regression (6) regresses monthly market star on the fund's 1-month lagged unexpected flow, Regression (7) regresses monthly market dog on the fund's 1-month lagged flow, Regression (8) regresses monthly market dog on the fund's 1-month expected lagged flow and Regression (9) regresses monthly market dog on the fund's 1-month lagged unexpected flow. All regressions have a time (year) and entity (fund) fixed effect and control for the fund's size (fund's average past year TNA), the fund's age (amount of months between time t and the date when the fund was offered for the first time) and the fund's style (fund's Lipper Objective Code). The standard error is clustered at the fund level.

	Dependent variable:								
	Market Rank (Alpha)			Market Star (Alpha)			Market Dog (Alpha)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Lag Flow	0.426***			0.173***			−0.158***		
	(0.029)			(0.018)			(0.017)		
Lag Exp. Flow		4.454***			1.670***			−1.647***	
		(0.208)			(0.134)			(0.129)	
Lag Unexp. Flow			0.248***			0.109***			−0.091***
			(0.021)			(0.014)			(0.013)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Entity FE	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund
Time FE	Year	Year	Year	Year	Year	Year	Year	Year	Year
Stand. Error	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.
Observations	146,117	144,901	144,901	146,117	144,901	144,901	146,117	144,901	144,901
Adjusted R^2	0.137	0.158	0.134	0.107	0.114	0.106	0.138	0.143	0.138

Note: .p<0.1;*p<0.05;**p<0.01;***p<0.001

Appendix 7.5: Family Smart Money Regression using Alpha

Table 18: Appendix: Family Smart Money Regression using Alpha

Table 18 shows the results of the regression within the family described in equation 32 using fund's alpha. Regression (1) regresses monthly fund's family rank on the fund's 1-month lagged flow, Regression (2) regresses monthly fund's family rank on the fund's 1-month lagged expected flow, Regression (3) regresses monthly fund's family rank on the fund's 1-month lagged unexpected flow, Regression (4) regresses monthly family star on the fund's 1-month lagged flow, Regression (5) regresses monthly family star on the fund's 1-month lagged expected flow, Regression (6) regresses monthly family star on the fund's 1-month lagged unexpected flow, Regression (7) regresses monthly family dog on the fund's 1-month lagged flow, Regression (8) regresses monthly family dog on the fund's 1-month expected lagged flow and Regression (9) regresses monthly family dog on the fund's 1-month lagged unexpected flow. All regressions have a time (year) and entity (fund) fixed effect and control for the fund's size (fund's average past year TNA), the fund's age (amount of months between time t and the date when the fund was offered for the first time) and the fund's style (fund's Lipper Objective Code). The standard error is clustered at the fund level.

	Dependent variable:								
	Family Rank (Alpha)			Family Star (Alpha)			Family Dog (Alpha)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Lag Flow	0.312***			0.168***			−0.117***		
	(0.034)			(0.026)			(0.021)		
Lag Exp. Flow		3.180***			1.429***			−1.223***	
		(0.247)			(0.185)			(0.158)	
Lag Unexp. Flow			0.184***			0.113***			−0.067***
			(0.025)			(0.021)			(0.017)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Entity FE	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F
Time FE	Year	Year	Year	Year	Year	Year	Year	Year	Year
Stand. Error	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.
Observations	89,211	88,570	88,570	89,211	88,570	88,570	89,211	88,570	88,570
Adjusted R^2	0.126	0.134	0.125	0.104	0.106	0.104	0.109	0.112	0.110

Note: Fund & Family (F & F) .p<0.1;*p<0.05;**p<0.01;***p<0.001

Appendix 8: Robustness Check

Appendix 8.1: Robustness Check I Star and Dog Threshold

Table 19: Robustness Check I: Star and Dog Threshold (Market Regression)

Table 19 shows the results of the market-wide regression described in equation 14 using, however, only the market star and dog variables with two different thresholds (best and worst 9% and 1% instead of previously 5%). Regression (1), (2), (3), (7), (8) and (9) use a threshold of 9%, which means that the star and dog funds are the best, respective worst, 9% performing funds. Regression (4), (5), (6), (10), (11) and (12) use a threshold of 1%, which means that the star and dog funds are the best, respective worst, 1% performing funds. For all 4 alternative variables, regressions are performed using the fund, expected and unexpected flow. All regressions have a time (year) and entity (fund) fixed effect and control for the fund's size (fund's average past year TNA), the fund's age (amount of months between time t and the date when the fund was offered for the first time) and the fund's style (fund's Lipper Objective Code). The standard error is clustered at the fund level.

	Dependent variable:											
	Market Star (9%)			Market Star (1%)			Market Dog (9%)			Market Dog (1%)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Lag Flow	0.057***			0.009			−0.071***			−0.014***		
	(0.017)			(0.006)			(0.016)			(0.005)		
Lag Exp. Flow		−0.276***			−0.100***			0.148*			−0.064**	
		(0.081)			(0.030)			(0.082)			(0.026)	
Lag Unexp. Flow			0.070***			0.013**			−0.078***			−0.012**
			(0.018)			(0.006)			(0.016)			(0.005)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Entity FE	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund
Time FE	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year
Stand. Error	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.
Observations	146,117	144,901	144,901	146,117	144,901	144,901	146,117	144,901	144,901	146,117	144,901	144,901
Adjusted R^2	0.055	0.055	0.055	0.038	0.038	0.038	0.083	0.083	0.083	0.070	0.070	0.070

Note: .p<0.1;*p<0.05;**p<0.01;***p<0.001

Table 20: Robustness Check I: Star and Dog Threshold (Family Regression)

Table 20 shows the results of the family regression described in equation 15 using, however, only the family star and dog variables with two different thresholds (best and worst 9% and 1% instead of previously 5%). Regression (1), (2), (3), (7), (8) and (9) use a threshold of 9%, which means that the star and dog funds are the best, respective worst. 9% performing funds. Regression (4), (5), (6), (10), (11) and (12) use a threshold of 1%, which means that the star and dog funds are the best, respective worst, 1% performing funds. For all 4 alternative variables, regressions are performed using the fund, expected and unexpected flow. All regressions have a time (year) and entity (fund and family) fixed effect and control for the fund's size (fund's average past year TNA), the fund's age (amount of months between time t and the date when the fund was offered for the first time) and the fund's style (fund's Lipper Objective Code). The standard error is clustered at the fund level.

	Dependent variable:											
	Family Star (9%)			Family Star (1%)			Family Dog (9%)			Family Dog (1%)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Lag Flow	−0.019			−0.016			0.047**			0.022		
	(0.022)			(0.015)			(0.022)			(0.016)		
Lag Exp. Flow		−0.411***			−0.296***			0.314***			0.147*	
		(0.109)			(0.081)			(0.121)			(0.087)	
Lag Unexp. Flow			−0.002			−0.003			0.039*			0.020
			(0.022)			(0.015)			(0.022)			(0.016)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Entity FE	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F
Time FE	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year
Stand. Error	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.
Observations	89,211	88,570	88,570	89,211	88,570	88,570	89,211	88,570	88,570	89,211	88,570	88,570
Adjusted R^2	0.057	0.058	0.057	0.064	0.065	0.065	0.082	0.082	0.082	0.093	0.093	0.093

Note: Fund & Family (F & F) .p<0.1;*p<0.05;**p<0.01;***p<0.001

Appendix 8.2: Robustness Check II Focus on the 10% biggest families

Table 21: Robustness Check II: 10% Biggest Families

Table 21 shows the results of the family regression as described in equation 15. This table reproduces the exact same regressions as shown in Table 5 in Chapter 4.3.1 but with a different data set. The data set used in this table only contains the 10% biggest families. All regressions have a time (year) and entity (fund and family) fixed effect and control for the fund's size (fund's average past year TNA), the fund's age (amount of months between time t and the date when the fund was offered for the first time) and the fund's style (fund's Lipper Objective Code). The standard error is clustered at the fund level.

	Dependent variable:								
	Family Rank			Family Star			Family Dog		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Lag Flow	−0.045			−0.043**			0.003		
	(0.029)			(0.021)			(0.020)		
Lag Exp. Flow		−0.684***			−0.447***			0.304**	
		(0.151)			(0.107)			(0.130)	
Lag Unexp. Flow			−0.016			−0.024			−0.007
			(0.029)			(0.021)			(0.020)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Entity FE	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F
Time FE	Year	Year	Year	Year	Year	Year	Year	Year	Year
Stand. Error	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.
Observations	55,508	55,105	55,105	55,508	55,105	55,105	55,508	55,105	55,105
Adjusted R^2	0.029	0.029	0.028	0.052	0.053	0.053	0.087	0.087	0.086

Note: Fund & Family (F & F) .p<0.1;*p<0.05;**p<0.01;***p<0.001

Appendix 8.3: Robustness Check III Decile Regressions

Table 22: Robustness Check III: Smart or Dumb Money (Fund Flow Decile)?
Table 22 shows the results of the fund flow market decile regression. This table uses the same methodology as the quintile regression described in Equation 16 in Chapter 3.3 using, however, deciles instead of quintiles. The Market Deciles 1 to 10 are calculated in the similar way as Market Quintiles 1 to 5, as described in Equation 12 in Chapter 3.2.5, but using 10 instead of 5 groups to classify funds based on their monthly performance. All regressions have a time (year) and entity (fund) fixed effect and control for the fund's size (fund's average past year TNA), the fund's age (amount of months between time t and the date when the fund was offered for the first time) and the fund's style (fund's Lipper Objective Code). The standard error is clustered at the fund level.

	Dependent variable:									
	Mark. Dec. 1	Mark. Dec. 2	Mark. Dec. 3	Mark. Dec. 4	Mark. Dec. 5	Mark. Dec. 6	Mark. Dec. 7	Mark. Dec. 8	Mark. Dec. 9	Mark. Dec. 10
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Lag Flow	0.061^{***}	0.015	0.013	−0.010	0.002	−0.007	0.013	−0.021	0.009	-0.075^{***}
	(0.018)	(0.016)	(0.014)	(0.014)	(0.016)	(0.015)	(0.017)	(0.015)	(0.016)	(0.016)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Entity FE	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund
Time FE	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year
Stand. Error	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.
Observations	146,117	146,117	146,117	146,117	146,117	146,117	146,117	146,117	146,117	146,117
Adjusted R^2	0.054	0.010	0.004	0.012	0.019	0.019	0.012	0.008	0.014	0.083

Note: .p<0.1;*p<0.05;**p<0.01;***p<0.001

Table 23: Robustness Check III: Market Decile Regression (Expected Flow)
Table 23 shows the results of the expected flow market decile regression. This table uses the same methodology as the quintile regression described in Equation 16 in Chapter 3.3 using, however, deciles instead of quintiles. The Market Deciles 1 to 10 are calculated in the similar way as Market Quintiles 1 to 5, as described in Equation 12 in Chapter 3.2.5, but using 10 instead of 5 groups to classify funds based on their monthly performance. All regressions have a time (year) and entity (fund) fixed effect and control for the fund's size (fund's average past year TNA), the fund's age (amount of months between time t and the date when the fund was offered for the first time) and the fund's style (fund's Lipper Objective Code). The standard error is clustered at the fund level.

	Dependent variable:									
	Mark. Dec. 1	Mark. Dec. 2	Mark. Dec. 3	Mark. Dec. 4	Mark. Dec. 5	Mark. Dec. 6	Mark. Dec. 7	Mark. Dec. 8	Mark. Dec. 9	Mark. Dec. 10
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Lag Exp. Flow	-0.255^{***}	-0.261^{***}	−0.121	−0.054	0.039	0.029	0.038	0.087	0.303^{***}	0.196^{**}
	(0.086)	(0.082)	(0.079)	(0.075)	(0.077)	(0.085)	(0.073)	(0.080)	(0.080)	(0.085)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Entity FE	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund
Time FE	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year
Stand. Error	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.
Observations	144,901	144,901	144,901	144,901	144,901	144,901	144,901	144,901	144,901	144,901
Adjusted R^2	0.055	0.010	0.004	0.013	0.019	0.019	0.013	0.008	0.014	0.082

Note: .p<0.1;*p<0.05;**p<0.01;***p<0.001

Table 24: Robustness Check III: Market Decile Regression (Unexpected Flow)
Table 24 shows the results of the unexpected flow market decile regression. This table uses the same methodology as the quintile regression described in Equation 16 in Chapter 3.3 using, however, deciles instead of quintiles. The Market Deciles 1 to 10 are calculated in the similar way as Market Quintiles 1 to 5, as described in Equation 12 in Chapter 3.2.5, but using 10 instead of 5 groups to classify funds based on their monthly performance. All regressions have a time (year) and entity (fund) fixed effect and control for the fund's size (fund's average past year TNA), the fund's age (amount of months between time t and the date when the fund was offered for the first time) and the fund's style (fund's Lipper Objective Code). The standard error is clustered at the fund level.

	Dependent variable:									
	Mark. Dec. 1	Mark. Dec. 2	Mark. Dec. 3	Mark. Dec. 4	Mark. Dec. 5	Mark. Dec. 6	Mark. Dec. 7	Mark. Dec. 8	Mark. Dec. 9	Mark. Dec. 10
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Lag Unexp. Flow	0.074***	0.025	0.018	−0.007	0.002	−0.007	0.010	−0.025*	−0.003	−0.085***
	(0.018)	(0.017)	(0.015)	(0.014)	(0.016)	(0.015)	(0.017)	(0.015)	(0.016)	(0.016)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Entity FE	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund	Fund
Time FE	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year
Stand. Error	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.
Observations	144,901	144,901	144,901	144,901	144,901	144,901	144,901	144,901	144,901	144,901
Adjusted R^2	0.055	0.010	0.004	0.013	0.019	0.019	0.013	0.008	0.014	0.082

Note: .p<0.1;*p<0.05;**p<0.01;***p<0.001

Table 25: Robustness Check III: Family Decile Regression (Fund Flow)
Table 25 shows the results of the fund flow family decile regression. This table uses the same methodology as the quintile regression described in Equation 17 in Chapter 3.3 using, however, deciles instead of quintiles. The Family Deciles 1 to 10 are calculated in the similar way as Family Quintiles 1 to 5, as described in Equation 13 in Chapter 3.2.5, but using 10 instead of 5 groups to classify funds based on their monthly performance. All regressions have a time (year) and entity (fund and family) fixed effect and control for the fund's size (fund's average past year TNA), the fund's age (amount of months between time t and the date when the fund was offered for the first time) and the fund's style (fund's Lipper Objective Code). The standard error is clustered at the fund level.

	Dependent variable:									
	Fam. Dec. 1	Fam. Dec. 2	Fam. Dec. 3	Fam. Dec. 4	Fam. Dec. 5	Fam. Dec. 6	Fam. Dec. 7	Fam. Dec. 8	Fam. Dec. 9	Fam. Dec. 10
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Lag Flow	−0.027	−0.029	−0.007	−0.009	−0.016	0.005	0.030	−0.009	0.013	0.049**
	(0.023)	(0.019)	(0.021)	(0.018)	(0.019)	(0.020)	(0.021)	(0.018)	(0.020)	(0.023)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Entity FE	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F
Time FE	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year
Stand. Error	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.
Observations	89,211	89,211	89,211	89,211	89,211	89,211	89,211	89,211	89,211	89,211
Adjusted R^2	0.061	0.011	0.008	0.014	0.022	0.024	0.015	0.012	0.016	0.083

Note: Fund & Family (F & F) .p<0.1;*p<0.05;**p<0.01;***p<0.001

Table 26: Robustness Check III: Family Decile Regression (Expected Flow)
Table 26 shows the results of the expected flow family decile regression. This table uses the same methodology as the quintile regression described in Equation 17 in Chapter 3.3 using, however, deciles instead of quintiles. The Family Deciles 1 to 10 are calculated in the similar way as Family Quintiles 1 to 5, as described in Equation 13 in Chapter 3.2.5, but using 10 instead of 5 groups to classify funds based on their monthly performance. All regressions have a time (year) and entity (fund and family) fixed effect and control for the fund's size (fund's average past year TNA), the fund's age (amount of months between time t and the date when the fund was offered for the first time) and the fund's style (fund's Lipper Objective Code). The standard error is clustered at the fund level.

	Dependent variable:									
	Fam. Dec. 1	Fam. Dec. 2	Fam. Dec. 3	Fam. Dec. 4	Fam. Dec. 5	Fam. Dec. 6	Fam. Dec. 7	Fam. Dec. 8	Fam. Dec. 9	Fam. Dec. 10
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Lag Exp. Flow	-0.433^{***}	-0.249^{**}	-0.080	-0.085	0.192^{*}	0.129	0.080	0.120	-0.002	0.328^{***}
	(0.118)	(0.096)	(0.104)	(0.110)	(0.102)	(0.097)	(0.106)	(0.108)	(0.104)	(0.122)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Entity FE	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F
Time FE	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year
Stand. Error	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.
Observations	88,570	88,570	88,570	88,570	88,570	88,570	88,570	88,570	88,570	88,570
Adjusted R^2	0.062	0.011	0.008	0.014	0.023	0.023	0.015	0.012	0.016	0.082

Note: Fund & Family (F & F) .p<0.1;*p<0.05;**p<0.01;***p<0.001

Table 27: Robustness Check III: Family Decile Regression (Unexpected Flow)
Table 27 shows the results of the unexpected flow family decile regression. This table uses the same methodology as the quintile regression described in Equation 17 in Chapter 3.3 using, however, deciles instead of quintiles. The Family Deciles 1 to 10 are calculated in the similar way as Family Quintiles 1 to 5, as described in Equation 13 in Chapter 3.2.5, but using 10 instead of 5 groups to classify funds based on their monthly performance. All regressions have a time (year) and entity (fund and family) fixed effect and control for the fund's size (fund's average past year TNA), the fund's age (amount of months between time t and the date when the fund was offered for the first time) and the fund's style (fund's Lipper Objective Code). The standard error is clustered at the fund level.

	Dependent variable:									
	Fam. Dec. 1	Fam. Dec. 2	Fam. Dec. 3	Fam. Dec. 4	Fam. Dec. 5	Fam. Dec. 6	Fam. Dec. 7	Fam. Dec. 8	Fam. Dec. 9	Fam. Dec. 10
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Lag Unexp. Flow	-0.009	-0.017	-0.005	-0.004	-0.024	0.001	0.024	-0.017	0.011	0.040^{*}
	(0.023)	(0.019)	(0.021)	(0.018)	(0.018)	(0.020)	(0.021)	(0.018)	(0.021)	(0.022)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Entity FE	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F	F & F
Time FE	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year
Stand. Error	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.	Clust.
Observations	88,570	88,570	88,570	88,570	88,570	88,570	88,570	88,570	88,570	88,570
Adjusted R^2	0.061	0.011	0.008	0.014	0.023	0.023	0.015	0.012	0.016	0.082

Note: Fund & Family (F & F) .p<0.1;*p<0.05;**p<0.01;***p<0.001

References

Ammann, M., Bauer, C., Fischer, S., and Müller, P. (2019). The impact of the Morningstar Sustainability Rating on mutual fund flows. *European Financial Management*, 25(3):520–553.

Bailey, W., Kumar, A., and Ng, D. (2011). Behavioral Biases of Mutual Fund Investors. *Journal of Financial Economics*, 102(1):1–27.

Barber, B. M., Odean, T., and Zheng, L. (2005). Out of Sight, Out of Mind: The Effects of Expenses on Mutual Fund Flows. *Journal of Business*, 78(6):2095–2119.

Benson, K. L., Tang, G., and Tutticci, I. (2008). The Relevance of Family Characteristics to Individual Fund Flows:. *Australian Journal of Management*, 32(3):419–443.

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

Berk, J. B. and van Binsbergen, J. H. (2015). Measuring Skill in the Mutual Fund Industry. *Journal of Financial Economics*, 118(1):1–20.

Blake, C. R. and Morey, M. R. (2000). Morningstar Ratings and Mutual Fund Performance. *Journal of Financial and Quantitative Analysis*, 35(3):451–483.

Blume, M. E. (1998). An Anatomy of Morningstar Ratings. *Financial Analysts Journal*, 54(2):19–27.

Bollen, N. P. B. and Busse, J. A. (2005). Short-Term Persistence in Mutual Fund Performance. *The Review of Financial Studies*, 18(2):569–597.

Brown, D. P. and Wu, Y. (2016). Mutual Fund Flows and Cross-Fund Learning within Families. *The Journal of Finance*, 71(1):383–424.

Bu, Q. and Lacey, N. (2008). Do Mutual Funds Exhibit a Smart Money Effect? *Quarterly Journal of Finance and Accounting*, 47(1):53–68.

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance*, 52(1):57–82.

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

Choi, J. J. and Robertson, A. Z. (2020). What Matters to Individual Investors? Evidence from the Horse's Mouth. *The Journal of Finance*, 75(4):1965–2020.

Coval, J. and Stafford, E. (2007). Asset Fire Sales (and Purchases) in Equity Markets. *Journal of Financial Economics*, 86(2):479–512.

CRSP (2018). Survivor-Bias-Free US Mutual Fund Guide. Technical report, Cenrer for Research in Security Prices.

El Ghoul, S. and Karoui, A. (2020). Fund Names vs. Family Names: Implications for Mutual Fund Flows. *SSRN Electronic Journal*, pages 1–46.

Elton, E. J., Gruber, M. J., and Blake, C. R. (1996). The Persistence of Risk-Adjusted Mutual Fund Performance. *Source: The Journal of Business*, 69(2):133–157.

Elton, E. J., Gruber, M. J., and Blake, C. R. (2001). A First Look at the Accuracy of the CRSP Mutual Fund Database and a Comparison of the CRSP and Morningstar Mutual Fund Databases. *The Journal of Finance*, 56(6):2415–2430.

Elton, E. J., Gruber, M. J., and Blake, C. R. (2006). The Adequacy of Investment Choices Offered by 401(k) Plans. *Investments and Portfolio Performance*, pages 307–322.

Elton, E. J., Gruber, M. J., and Green, T. C. (2007). The Impact of Mutual Fund Family Membership on Investor Risk. *Source: The Journal of Financial and Quantitative Analysis*, 42(2):257–277.

Fama, E. F. and French, K. R. (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance*, 47(2):427–465.

Ferreira, M. A. and Ramos, S. B. (2009). Mutual Fund Industry Competition and Concentration: International Evidence. *SSRN Electronic Journal*, pages 1–54.

Foad, H. (2011). Familiarity Bias. In Baker, K. H. and Nofsinger, J. R., editors, *Behavioral Finance: Investors, Corporations, and Markets*, pages 277–294. John Wiley & Sons, Ltd.

Frazzini, A. and Lamont, O. A. (2008). Dumb Money: Mutual Fund Flows and the Cross-Section of Stock Returns. *Journal of Financial Economics*, 88(2):299–322.

Gallaher, S. T., Kaniel, R., and Starks, L. T. (2008). Advertising and Mutual Funds: From Families to Individual Funds. *SSRN Electronic Journal*, pages 1–48.

Gaspar, J.-M., Massa, M., and Matos, P. (2006). Favoritism in Mutual Fund Families? Evidence on Strategic Cross-Fund Subsidization. *The Journal of Finance*, 61(1):73–104.

Gerken, W. C., Starks, L. T., and Yates, M. (2018). The Importance of Family: The Role of Mutual Fund Family Reputation in Investment Decisions. *SSRN Electronic Journal*, pages 1–54.

Gruber, M. J. (1996). Another Puzzle: The Growth in Actively Managed Mutual Funds. *The Journal of Finance*, 51(3):783–810.

Guedj, I. and Papastaikoudi, J. (2003). Can Mutual Fund Families Affect the Performance of Their Funds? *SSRN Electronic Journal*, pages 1–44.

Hazenberg, J. J., Irek, F., Scheer, W. v. d., and Stefanova, M. (2015). The Lure of the Brand: Evidence from the European Mutual Fund Industry. *European Financial Management*, 21(5):867–904.

Hendricks, D., Patel, J., and Zeckhauser, R. (1993). Hot Hands in Mutual Funds: Short-Run Persistence of Relative Performance, 1974-1988. *Source: The Journal of Finance*, 48(1):93–130.

Heuer, J., Merkle, C., and Weber, M. (2017). Fooled by Randomness: Investor Perception of Fund Manager Skill. *Review of Finance*, 21(2):605–635.

Investment Company Institute (2021). 2021 Investment Company FACT BOOK: A Review of Trends and Activities in the Investment Company Industry. Technical report, Investment Company Institute.

Ippolito, R. A. (2015). Consumer Reaction to Measures of Poor Quality: Evidence from the Mutual Fund Industry. *The University of Chicago Press Journals*, 35(1):45–70.

Ivkovic, Z. and Weisbenner, S. (2005). Local Does as Local Is: Information Content of the Geography of Individual Investors' Common Stock Investments. *The Journal of Finance*, 60(1):267–306.

Jensen, M. C. (1967). The Performance Of Mutual Funds In The Period 1945-1964. *Journal of Finance*, 23(2):389–416.

Jiang, G. J. and Yuksel, H. Z. (2017). What Drives the "Smart-Money" Effect? Evidence from Investors' Money Flow to Mutual Fund Classes. *Journal of Empirical Finance*, 40:39–58.

Joo, H.-K. and Park, Y. K. (2011). Contribution of Star Funds to Fund Families: An Empirical Analysis of the Korean Fund Market. *Asia-Pacific Journal of Financial Studies*, 40(5):731–762.

Kaniel, R. and Parham, R. (2017). WSJ Category Kings – The Impact of Media Attention on Consumer and Mutual Fund Investment Decisions. *Journal of Financial Economics*, 123(2):337–356.

Kempf, A. and Ruenzi, S. (2008a). Family Matters: Rankings within Fund Families and Fund Inflows. *Journal of Business Finance and Accounting*, 35(1-2):177–199.

Kempf, A. and Ruenzi, S. (2008b). Tournaments in Mutual-Fund Families. *The Review of Financial Studies*, 21(2):1013–1036.

Keswani, A. and Stolin, D. (2008). Which Money Is Smart? Mutual Fund Buys and Sells of Individual and Institutional Investors. *The Journal of Finance*, 63(1):85–118.

Kronlund, M., Pool, V. K., Sialm, C., and Stefanescu, I. (2020). Out of Sight no More? The Effect of Fee Disclosures on 401(k) Investment Allocations. *SSRN Electronic Journal*, pages 1–78.

Lou, D. (2012). A Flow-Based Explanation for Return Predictability. *The Review of Financial Studies*, 25(12):3457–3489.

Massa, M. (2003). How do Family Strategies Affect Fund Performance? When Performance-Maximization is not the Only Game in Town. *Journal of Financial Economics*, 67(2):249–304.

Nanda, V., Wang, Z. J., and Zheng, L. (2004). Family Values and the Star Phenomenon: Strategies of Mutual Fund Families. *The Review of Financial Studies*, 17(3):667–698.

Reuter, J. and Zitzewitz, E. (2006). Do Ads Influence Editors? Advertising and Bias in the Financial Media. *The Quarterly Journal of Economics*, 121(1):197–227.

Roussanov, N. L., Ruan, H., and Wei, Y. M. (2020). Mutual Fund Flows and Performance in (Imperfectly) Rational Markets? *SSRN Electronic Journal*, pages 1–48.

Salganik, G. (2013). The "Smart Money" Effect: Retail versus Institutional Mutual Funds. *The Journal of Behavioral Finance & Economics*, 3(1):21–71.

Sapp, T. and Tiwarsi, A. (2004). Does Stock Return Momentum Explain the “Smart Money” Effect? *The Journal of Finance*, 59(6):2605–2622.

Sawicki, J. and Finn, F. (2002). Smart Money and Small Funds. *Journal of Business Finance & Accounting*, 29(5-6):825–846.

Seasholes, M. S. and Zhu, N. (2010). Individual Investors and Local Bias. *The Journal of Finance*, 65(5):1987–2010.

Sialm, C. and Tham, T. M. (2015). Spillover Effects in Mutual Fund Companies. *Management Science*, 62(5):1472–1486.

Simutin, M. (2013). Standing Out in the Fund Family: Deviation from a Family Portfolio Predicts Mutual Fund Performance. *SSRN Electronic Journal*, pages 1–49.

Sirri, E. R. and Tufano, P. (1998). Costly Search and Mutual Fund Flows. *The Journal of Finance*, 53(5):1589–1622.

Song, Y. (2020). The Mismatch Between Mutual Fund Scale and Skill. *The Journal of Finance*, 75(5):2555–2589.

Verbeek, M. and Huij, J. (2007). Spillover Effects of Marketing in Mutual Fund Families. *SSRN Electronic Journal*, pages 1–43.

Vidal-García, J. (2013). The Persistence of European Mutual Fund Performance. *Research in International Business and Finance*, 28(1):45–67.

Wei, A.-P., Chen, M.-L., and Peng, C.-L. E. (2011). The Advertising Spillover Effect: Implications for Mutual Fund Families. *Journal of Management*, 28(4):361–377.

Wermers, R. (2003). Is Money Really 'Smart'? New Evidence on the Relation between Mutual Fund Flows, Manager Behavior, and Performance Persistence. *SSRN Electronic Journal*, pages 1–53.

Zheng, L. (1999). Is Money Smart? A Study of Mutual Fund Investors' Fund Selection Ability. *The Journal of Finance*, 54(3):901–933.

Zheng, L. (2008). The Behavior of Mutual Fund Investors. In *Handbook of Financial Intermediation and Banking*, pages 259–283. Elsevier.