ARE INVESTORS BETTER SAFE THAN SORRY?

THE IMPACT OF EXTREME LOSSES IN THE RETURN DISTRIBUTION ON CAPITAL ALLOCATION IN ACTIVELY MANAGED EQUITY FUNDS

YLVALI BUSCH

GUSTAV GÖRANSSON

Bachelor Thesis

Stockholm School of Economics

2020

Are investors better safe than sorry? The impact of extreme losses in the return distribution on capital allocation in actively managed equity funds

Abstract:

In this paper, we show that capital flow to actively managed equity funds is dependent on past extreme negative return states of the fund. Specifically, we examine how an extreme negative monthly payoff impacts the investment flow of actively managed equity funds in the following year, adjusting for past performance and other fund characteristics. Our results indicate that investors make their investment decisions in line with one of the predictions of cumulative prospect theory, namely that investors place excess weight on tail events. As a result, they are less willing to direct investments into funds with an extreme negative payoff in their historical returns as they overweight the probability of similar extreme events in the future. Furthermore, we examine the practical implications of this investor behavior, discussing the impact on fund managers and their portfolio strategy. Our results are robust even when controlling for factors such as historical performance, volatility, fund fees, fund size and company size.

Keywords:

Mutual funds, actively managed equity funds, fund flows, extreme payoffs, prospect theory

Authors:

Ylvali Busch (24055)

Gustav Göransson (24281)

Tutor:

Adam Altmejd, Postdoctoral Fellow, Swedish House of Finance

Examiner:

Adrien D’Avernas, Assistant Professor, Swedish House of Finance

Bachelor Thesis
Bachelor Program in Business and Economics
Stockholm School of Economics

This page is intentionally left blank

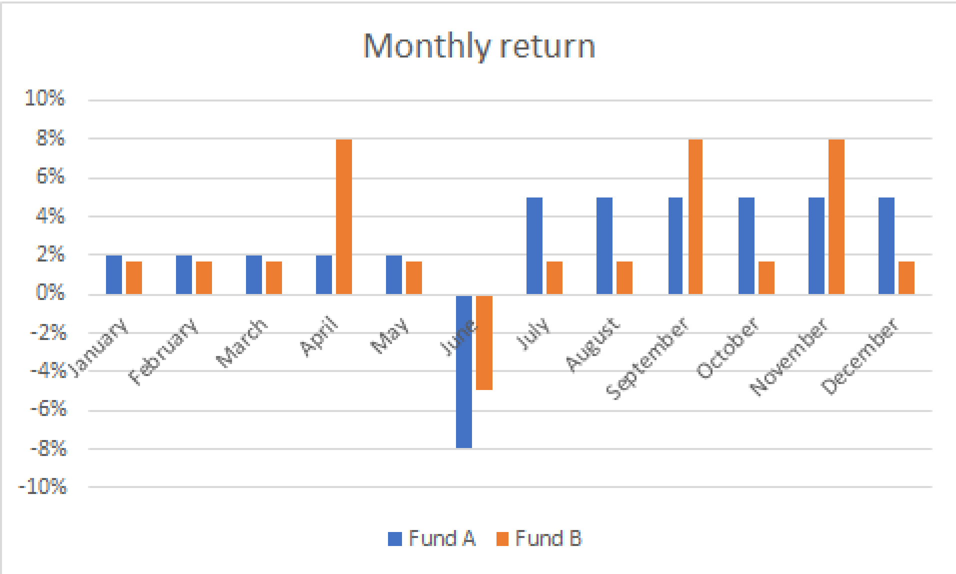
1. Introduction

This paper examines how the demand for actively managed equity funds (a subset of mutual funds) is affected by the fund return distribution, showing that a trailing extreme negative return on record impacts capital flow to the fund. The market dynamics behind the demand for actively managed equity funds is important to understand, as fund flows can impact both incentives for the fund managers as well as asset prices (Coval and Stafford, 2005; Wermers, 2005; Sensoy, 2008). At the same time, retail investors hold 89 percent of net assets of mutual funds (Investment Company Institute, 2019), making the demand for funds subject to individual investor biases and preferences. Such biases have previously been shown to deleteriously affect the financial well-being of retail investors (Barber and Odean, 2012; Fred van Raaij, 2016). As a result, understanding such biases and preferences becomes key in order to protect the financial well-being of retail investors by improving the market structure.

Within the previous research on investor biases and preferences, cumulative prospect theory is a central model that can be used to understand the behavior of retail investors. According to cumulative prospect theory, individual investors overweight *extreme* outcomes in their decision-making. More specifically, individual investors tend to overweight the probability of tail events, i.e. rare and high-impact events (Barberis, 2013). This model has shown a higher accuracy in predicting investor behavior compared to the expected utility framework (Tversky and Kahneman, 1981, 1992; Tversky and Fox, 1995; Wu and Gonzalez, 1996; Stott, 2006; Barberis and Huang, 2008; Abdellaoui, Bleichrodt and Kammoun, 2013; Barberis, 2013). Akbas and Genc (2020) (hereafter AG, 2020) contribute to the understanding of how demand for mutual funds relates to distributional features beyond mean performance by exploring how extreme positive returns impact fund flow. However, the other side of the spectrum has been left unexplored, namely how fund demand is affected by extreme negative returns. Thus, our work is an extension of AG (2020), as we explore the impact of extreme negative returns on fund flows while AG (2020) explore extreme positive payoffs. Prospect theory predicts that both positive and negative extreme tail events in the return distribution should impact investor behavior (Kahneman and Tversky, 1979, 1992), hence our research has theoretical relevance.

Specifically, we explore how a single low-payoff state impacts fund flow, adjusting for past performance and other fund characteristics. Evidence by AG (2020) suggests that investors prefer funds that have an extreme positive monthly payoff (12-month trailing) on record. We will from now on refer to this 12-month trailing highest monthly return as MAX, just as in AG (2020). Adjusting for the cumulative return over a twelve-month period, we aim to determine what impact the twelve-month trailing style-adjusted minimum monthly return (hereafter, MIN) has on fund flow. Figure 1 illustrates the MIN graphically by comparing two funds with identical cumulative return and volatility (see Appendix I), but different MINs.

Figure 1. Comparing the style-adjusted monthly returns of two fictitious funds over twelve months with identical cumulative return and volatility, fund A has a more extreme MIN (June).



Our results indicate that MIN impacts fund flow to actively managed equity funds, meaning investors are subject to a behavior that does not correspond to expected utility theory. Instead, investors make their decisions in line with one of the concepts within prospect theory, placing excess weight on low probability events. As a result, they are less willing to direct investments into funds with an extreme negative payoff on record as they overweight the probability of similar extreme events in the future. These results are robust even when controlling for factors such as historical performance, volatility, management fee, fund size and company size. Interestingly, our findings indicate that MAX has a larger impact than MIN on future fund flow.

We contribute to the understanding of investor preferences in two ways. First, our research expands the theoretical understanding of investor behavior in relation to tail events for mutual funds, which is a relatively unexplored product in this context compared to other financial products (AG, 2020). Second, our research contributes to the understanding of investor behavior in relation to assets with distributional features of crash-like returns rather than lottery-like returns. Furthermore, we examine the practical implications of this investor behavior, discussing the impact on fund managers and their portfolio strategy.

2. Literature Review

2.1 Investor psychology of tail events

Investor psychology of tail events builds on the notion that people tend to overestimate and overweight the occurrences of low-probability and high-impact events, in line with cumulative prospect theory developed by Tversky and Kahneman (1992). This theory implies that a person will both overestimate the likelihood of tail events, as well as put more weight to the event in their decision-making, than what would be expected from a utility framework (Barberis, 2013). However, in this study we will only be able to observe the aggregate effects of these two concepts and not the separate effects, as they are hard to distinguish between in a non-laboratory setting.

Previous empirical research indicates that tail-events are indeed overestimated and overweighted in decision making, resulting in an excess demand for lottery-like assets as a historical right-tail outcome yields an expectation of similar upsides in the future (Brunnermeier, Gollier and Parker, 2007; Barberis and Huang, 2008; Eraker and Ready, 2015; AG, 2020). For instance, research by Barberis and Huang (2008) indicates that the skewness in the distribution of asset returns can be priced. Positively skewed stocks will be overpriced and earn a lower average return. This is due to the fact that the (unlikely) possibility of the stock experiencing a right-tail outcome, yielding the investor an extreme return, is overweighted in the decision-making. Hence, investors are willing to accept a higher price and lower return. The evidence for the opposite, that investors would shy away from negatively skewed stocks could be found by looking at the high historical U.S. equity premium, also known as the "equity premium puzzle". According to the probability weighting, the negative skewness of the aggregated stock market, due to historical large crashes, would imply that investors overweight these negative tail events, and therefore require a higher equity premium (De Giorgi and Legg, 2012).

Although the evidence supports the phenomenon of investors overestimating and overweighting the likelihood of tail events, there have also been instances where the likelihood of such events might have been underestimated. An example of this could be the 2008 financial crisis (Barberis, 2013). A central concept to why some tail events might be overestimated while others might be underestimated, is the "availability heuristics". The availability heuristics states that investors value the probability of an event by how easily such an event can be recalled (Barberis, 2013).

2.2 Investor psychology of tail events applied to mutual funds

Even though most mutual fund assets are held by individual investors, the understanding of how distributional features impact demand is low for mutual funds compared to other financial products, such as stocks and options (AG, 2020). Research by AG (2020) suggests that investors prefer funds that have an extreme positive monthly payoff (12-month trailing) on record in line with cumulative prospect theory (preference hypothesis). Hence, AG's (2020) results support that investor preferences for gambling-like payoffs are also valid for mutual funds. To corroborate this, AG (2020) explore two alternative explanations to the effects of MAX on fund flow, namely that MAX is a predictor of future fund performance and that MAX impacts the fund visibility.

Concerning the first alternative explanation, if MAX is a predictor of future returns, investors would simply pick these funds to maximize return, in line with expected utility theory. However, AG (2020) do not find support that a trading strategy based on MAX would yield superior returns.

Concerning the second alternative explanation, AG (2020) acknowledge that fund visibility (fund visibility hypothesis) can have an impact on the relationship between MAX and fund flow but argue that increased visibility could not account for the entire relationship. It is reasonable to believe that an extreme payoff would yield significant attention to a fund. However, this would only contribute to including the fund in the option space for investors. The actual choice would still be subject to preference. AG (2020) find evidence to support that the relationship between MAX and fund flow only holds in equity funds with a high degree of active management, suggesting that it is the risk-seeking investors that prefer the lottery-like payoffs. Furthermore, AG (2020) investigate proxies for fund visibility and find that none of these proxies substantially alter the relationship between MAX and fund flow. Finally, when the preference hypothesis predicts that reoccurring MAX in historical returns would enhance investor expectations of similar returns in the future, the visibility hypothesis predicts the opposite. If a fund has recently experienced a high MAX, additional high MAXs should not have the same effect, as the fund is already in the investor's option space. Consistent with the preference hypothesis and not the visibility hypothesis, AG (2020) find evidence that the relationship between MAX and fund flow is stronger for funds that have experienced high MAXs in the past.

Although AG (2020) use several types of mutual funds when they perform additional tests, their main experiment is based on actively managed equity funds. The reason being that the effect is only pronounced for this specific type of mutual fund. AG (2020) suggest that this distinction could be because actively managed funds possess a relatively flexible investment mandate, enhancing the possibility to make risky investments and generate the volatility required to obtain tail events. Hence attracting more risk-seeking investors.

3. Theoretical framework

3.1 Cumulative prospect theory

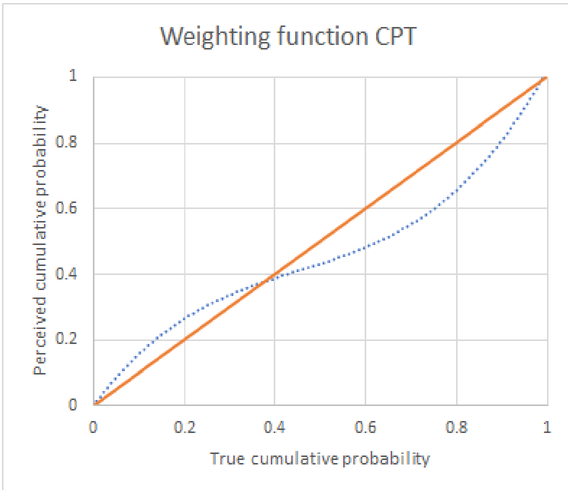
Cumulative prospect theory[1] was developed by Kahneman and Tversky (1992) as a response to the predictive shortcomings of expected utility theory. Instead of the notion that investors are rational in their decision-making from an expected utility point of view, prospect theory predicts that preferences for gains and losses are asymmetrical. Furthermore, the theory implies that an individual will overweight the tails of any distribution that is considered in a decision-making process. Prospect theory can be divided into two main concepts. The first concept is loss aversion, where investors are predicted to prefer avoiding losses compared to receiving a tantamount gain, having a concave utility function when faced with a gain, but a convex utility function when faced with a loss. For this research, the second main concept is however more relevant, concerning how individuals overweight extreme events[2].

Kahneman and Tversky (1979) show the overweighting of small probabilities mathematically with a stated probability p (the observed probability) along with the function $\pi(p)$ (the perceived probability). This definition was originally formulated for prospect theory, but still holds for cumulative prospect theory. For small p, their empirical research indicates that this function is subadditive:

$$0 < r < 1$$
$$\pi(rp) > r * \pi(p)$$

The graphical representation[3] of cumulative prospect theory (Tversky and Kahneman, 1992) shows the overweighting of stated or observed probabilities.

Figure 2. Investor cumulative density function in cumulative prospect theory. Blue dotted line shows the discrepancy of investor's weighting.



[1] Our theoretical framework is based on cumulative prospect theory (Kahneman and Tversky, 1992) rather than the original prospect theory (Kahneman and Tversky, 1979) due to the wide adoption of cumulative prospect theory in research where probability weighting is applied. This stems from the fact that cumulative prospect theory addresses important limitations of the original probability weighting function as it predicts that individuals will only overweight extreme low-probability outcomes, instead of all low-probability outcomes (Barberis, 2013).

[2] The empirical research done by Kahneman and Tversky (1979) relies on the subjects being in the gain domain or the loss domain. In our dataset, it is impossible to know which of these domains that the investor is in, and it is therefore hard to test the loss aversion.

[3] This function is slightly different for gains and losses, but this difference is in our case neglectable

3.2 MAX and MIN

The MAX variable was chosen by AG (2020) as existing literature suggests that investors overweight positive extreme states in their decision-making, rather than measures of the entire distribution (Brunnermeier, Gollier and Parker, 2007; Barberis and Huang, 2008). Furthermore, for investors to consider MAX in their decision-making, MAX must be accessible to investors (according to the principle of availability heuristics). Both in terms of the actual information, but also in terms of investor understanding of MAX. Many financial sites report historical lowest and highest monthly returns from the recent year. AG (2020) therefore argue that the extensive adoption of online platforms to retrieve financial information has made MAX available to investors. Furthermore, AG (2020) point to the fact that MAX is easier for retail investors to grasp compared to more complex measures such as skewness.

The main independent variable of interest in this study is MIN, which is used to measure a negative extreme return state in the monthly style-adjusted return over the past 12-months. This stems from a similar reasoning as in AG's (2020) choice of MAX, where the objective is to measure an extreme state rather than an entire distribution. Furthermore, we draw on the same logic as AG (2020) concerning the availability of MIN. Since financial sites report historical lowest and highest monthly returns, MIN should be available to investors in the same way as MAX.

4. Variables and Data sample

4.1 Variable definitions

For the time-constrained reader, we will summarize the differences between the variables used in this study and by AG (2020). First, we have added the MIN variable, as this is our extension of their study. There is a total of two variables that are unavailable to us that were used by AG (2020). The first variable is the total load fee, further explained below. Instead of total load fee, we use management fee, also further explained below. The second variable is the continuous performance ranking, used in the quadratic regression by AG (2020). We do not replace this variable, instead we use the discrete performance ranking that was also used by AG (2020), further explained below. All other variables are according to the procedures described by AG (2020). In addition, for the time-constrained reader, the variables used are summarized in table 1.

The dependent variable of interest is the quarterly fund flow as a percentage of current total net assets, FLOW, as defined by Sirri & Tuffano (1998):

$$FLOW_{i,t} = \frac{TNA_{i,t} - TNA_{i,t-1} * (1 + R_{i,t})}{TNA_{i,t-1}}$$

As fund flow is not given in the database that we use (CRSP database for mutual funds), it is calculated by using the difference in TNA, and subtracting the growth resulting from the fund return. The above formula calculates the growth of TNA accrued to the fund flow under the assumption that all capital flows occur at the end of the quarter. By applying the analysis at the quarterly level, comparisons with prior research is facilitated as this is the common level of analysis (AG, 2020).

The main independent variable in this paper, MIN, is the *minimum* monthly style-adjusted return over the past 12-months.

$$MIN_{i,t} = minimum(ADJ\_R_{i,t-11}, ADJ\_R_{i,t-10}, \dots, ADJ\_R_{i,t})$$

The independent variable MAX, the main variable of interest in the article by AG (2020), is the *maximum* monthly style-adjusted return over the past 12-months.

$$MAX_{i,t} = maximum(ADJ\_R_{i,t-11}, ADJ\_R_{i,t-10}, \dots, ADJ\_R_{i,t})$$

Both MIN and MAX are based on the style-adjusted return. Therefore, we construct a help-variable (ADJ_R) by subtracting the average return within each style from each fund's return every month. The style-adjustment of return is performed to account for the fact that funds within the same fund style often have correlated returns. This correlation exists because different styles restricts what type of stocks a fund can trade according to the style's investment mandate. Hence, a fund that belongs to a "winning"[4] style is also more likely to have superior returns. Furthermore, investment decisions are likely to be based on fund style, and different styles can periodically attract more attention in the media, as suggested by the articles Mullainathan, (2002), Barberis and Shleifer (2003) and Pomorski (2011). Hence, to differentiate the attraction of different styles from the effects of MIN (or MAX) in our study, style-adjusted return is appropriate.

Apart from the above, we also use several control variables that are probably more familiar to the reader. FUND_TNA is the aggregated total net assets for every share class in a fund (identified by the CRSP variable crsp_cl_grp). COMP_TNA is the aggregated total net assets for every fund in the same management company (identified by the CRSP variable mgmt_cd). AGE is the number of months since the fund was first offered (identified by the CRSP variable first_offer_dt). Here, the date of first offer is used as a proxy for the fund incubation date.

[4] The "winning" style is a fund style where the comprising funds have performed better on an aggregate level compared to other fund styles.

AG (2020) use the two variables total load fee[5] and expense ratio[6] to control for fund fees in the regression. However, the data used to aggregate load fees was unavailable to us and we therefore use the management fee instead of total load fee. MGMT_FEE is the management fee charged by the fund company (identified by the CRSP variable mgmt_fee). The management fee is a percentage fee on capital allocated to the fund. However, we cannot draw any conclusions on whether management fee and total load fees could function as good substitutes in a regression setting. EXP_RATIO is the expense ratio of the fund (identified by the CRSP variable exp_ratio) and TURN_RATIO is the turnover ratio of the fund (identified by the CRSP variable turn_ratio). The expense ratio is the amount that shareholders pay for operating expenses over their total investment. Turnover ratio is defined as the minimum of aggregated sales or aggregated purchases of securities over average 12-month TNA.

VOL is the 12-month trailing annualized volatility of fund return:

$$VOL_{i,t} = \sqrt{\frac{\sum_{t=0}^{-11}(R_{i,t} - \bar{R})^2}{11}}$$

SKEW is the 12-month trailing skewness of fund return:

$$SKEW_{i,t} = \left(\frac{\frac{\sum_{t=0}^{-11}(R_{i,t} - \bar{R})^2}{12}}{\sqrt{\frac{\sum_{t=0}^{-11}(R_{i,t} - \bar{R})^2}{11}}} \right)^3 = \frac{m_3}{s^3}$$

Where s is the sample standard deviation (also used for VOL) and $m_3$is the third central moment.

MIN and MAX values always represent the trailing 12-month's most extreme returns compared to similar funds[7]. Hence, to draw conclusions whether investors value the effect of extreme returns beyond the impact on aggregated returns, it is essential that we include a control variable for the fund's relative performance. Without this variable, the case might be that a high (low) extreme return simply reflects a high (low) aggregate return, rendering any regression results rather bland. A variable that allows the regression to capture the sensitivity of the flow due to the fund's relative performance is especially important, since the performance-flow relationship is well documented in the literature (Chevalier and Ellison, 1997; Brown, Harlow and Starks, 1996). In our model, we represent fund performance by grouping share classes together based on their relative adjusted performance

[5] AG (2020) use a weighted average of the front load and rear load fees to obtain this total load fee.
[6] Expense ratio is defined as the amount that shareholders pay for operating expenses over their total investment.

[7] To decide which funds that are comparable we use the CRSP objective code, which classifies mutual funds into several fund categories, called styles. The styles are further explained in the Data filtering section.

each month. We perform the ranking process by calculating the adjusted return (ADJ_R) threshold values that correspond to the bottom and top quintile. We calculate this once every month for every fund style. Based on this, for every month, for every share class of all funds, we assign if it has performed according to the bottom quintile (LOW_PERF), top quintile (HIGH_PERF), or the three middle quantiles (MID_PERF). Share classes are only ranked within their respective fund style. This follows the same procedure as AG (2020)[8], developed by Sirri and Tuffani (1998).

Table 1. Summary of variables as a courtesy to the reader

Variable name	**Description**
FLOW	The growth rate of a fund due to investment flows.
MAX	The maximum monthly style-adjusted return over the past 12-months.
MIN	The minimum monthly style-adjusted return over the past 12-months.
FUND_TNA	The aggregated total net assets for every share class in a fund.
COMP_TNA	The aggregated total net assets for every fund in the same management company.
AGE	The number of months since the fund was first offered.
MGMT_FEE	The management fee charged by the fund company.
EXP_RATIO	The expense ratio of the fund.
TURN_RATIO	The turnover ratio of the fund.
VOL	The 12-month trailing annualized volatility of fund return.
SKEW	The 12-month trailing skewness of the fund return.
LOW_PERF	This variable indicates that this month's adjusted return belongs to the bottom quintile of the adjusted returns in that particular month.

[8] We have only done the piecewise linear regression and not the quadratic regression from AG (2020). This is because a quadratic regression would require a continuous performance ranking, instead of the one based on 3 intervals (low, mid and high). We have not had access to the processing power required to compute a continuous performance ranking.

MIDDLE_PERF	Indicates that this month's adjusted return belongs to the middle 3-quintiles (20-80 percentile) of the adjusted returns in that particular month.
HIGH_PERF	This variable indicates that this month's adjusted return belongs to the top quintile of the adjusted returns in that particular month.

4.2 Data filtering

The goal is to handle the data filtering in the exact same way as AG (2020). However, the data filtering is not described by AG (2020) at a level of granularity such that it would be possible for us to do an exact replication of their data filtering. We believe that our data filtering is similar enough to AG (2020) to enable us to compare our results with theirs. A full guide to our data filtering can be found in Appendix III.

The database used is the Wharton Research Data Services (WRDS) Center for Research in Security Prices (CRSP) survivor-bias free data on mutual funds. Our data sample covers actively managed domestic U.S. equity funds. Therefore, bond, balanced, international and sector funds are excluded by including funds denoted by CRSP objective codes EDCI, EDCM, EDCS, EDYB, EDYG or EDYI[9]. Each CRSP objective code corresponds to a so-called fund style. Funds with CRSP index flags, indicating that they are index funds, are excluded from our sample[10] (AG, 2020). Since the CRSP index fund flag is only available after 2001, we also remove funds with names indicating that they are index funds from the sample by targeting names commonly used by such funds[11] All monthly data of funds without a valid CRSP objective code are also eliminated from our sample. The same filtering process to achieve the desired fund types is also applied by AG (2020).

The data is concatenated from the CRSP fund summary source and the CRSP monthly return source in order to calculate all relevant variables. We delimit the time period to August 1998 and December 2016. The exact starting point of August 1998 is chosen as the central variable crsp_cl_grp, which is used to determine which share class belongs to which fund, is not reported before that date. Regarding missing values in the dataset, our guiding principle is to remove months with missing data, rather than eliminating entire share-classes or funds, like AG (2020). Months that

[9] A specification on the CRSP objective codes can be found here: http://www.crsp.org/products/documentation/crsp-style-code-0

[10] Funds with CRSP index flags (variable index_fund_flag) "B" (index-based fund) and "D" (index fund) have been excluded from our sample, but funds with index flag "E" (enhanced fund) have been reincluded.

[11] Funds with the following strings contained in the fund names have been excluded from our sample: "Index", "Idx", "Indx", "S&P 500", "Dow Jones" and "BARRA". Fund names containing "Enhanced" or "Enh" have been reincluded in our sample.

lack information regarding TNA, management fee, expense ratio or turnover ratio are eliminated, as this prohibits us from calculating necessary control variables in the regression. Months that lack fund codes and company codes are also removed, as the absence of these variables prohibits us from assigning the share class to a specific fund and management company. Concerning missing return data, our approach has been to make sure at least twelve consecutive months exist, so that the trailing MIN, MAX and returns can be calculated. Hence, portions of a time series for a share-class can be removed if there is not sufficient data to retrieve the twelve-month trailing variables. Furthermore, we eliminate erroneous values from expense ratio and management fee, where only values between 0 (0%) and 1 (100%) are allowed[12]. We eliminate data where turnover ratio displays values smaller than 0 (0%).

Together with Morningstar, CRSP has been the most widely used mutual fund database in recent research. One major difference between the two databases is that CRSP is free from traditional survivorship bias. However, Elton, Gruber and Blake (2001) show that CRSP has a form of survivorship bias which they call *omission bias*. The omission bias is a result of the return data being inconsistently reported. While CRSP reports monthly return data for some funds, it reports annual or no return data for other funds. As it turns out, the merger and liquidation rates are much lower for the funds that have monthly return data, implying that any study using monthly return data will experience a skewed sample. The skewed sample is a consequence of an understating of the proportion of mergers and liquidations. Elton, Gruber and Blake (2001) conclude that CRSP can in fact display similar problems to datasets with traditional survivorship bias, such as overstating the performance of funds. As AG (2020) point out, this skewness could introduce extreme values around merger-dates when FLOW is calculated. To mitigate such extreme values of FLOW, the bottom and top 1% tails of FLOW are eliminated. In addition, funds that have less than 1 year of reported returns are eliminated. This is done to mitigate any potential incubation bias, meaning that new funds attract more flow and outperform other funds in terms of risk-adjusted return (Evans, 2010). Both of these adjustments are also done in AG (2020). Furthermore, Elton, Gruber and Blake (2001) also find that CRSP has upward biased return data compared to Morningstar, which is especially problematic for older data and small funds. As a result, we eliminate all share classes with less than $500,000 in total net assets.

[12] AG (2020) have not described an exact way of filtering out erroneous values in the dataset, but they describe that erroneous values are filtered out.

An important thing to note is that the data in our sample is calculated on share class level and not on fund level. In the article by AG (2020), share class data was aggregated to fund level, one of the reasons being that fund load fee could be aggregated (estimated) over share classes, to achieve a total fund load factor that applies to both front and rear load fees[13]. However, this data has not been available[14] during our work, and our analysis is therefore done on a share class level. One argument for aggregating share classes to funds is the aesthetic aspect of achieving results on fund level rather than share class level, but we hope that the reader can sympathize with this decision.

Another difference between our study and AG (2020) is that they have used data from January 1992, whereas our dataset starts in August 1998. This is since the important variable CRSP share class code (CRSP variable crsp_cl_grp was unavailable before August 1998. We are unaware of how AG (2020) handled this issue, but it is possible that the CRSP database has changed since their study was done, as the dataset used in AG (2020) was first compiled in 2016, whereas our dataset is compiled in 2020. The monthly data is transformed into a quarterly data set, since the regression is done on quarterly data. The cleaned dataset contains 330,307 quarter observations from 10,620 share classes of 3,621 funds, which is reasonably close[15] to the cleaned dataset used by AG (2020).

4.3 Descriptive statistics

In table 3, we observe the means of the regression variables over years. We note the large change in AGE, FUND_TNA and COMP_TNA between 1999 and 2000, likely reflecting the effect of the dot com bubble.

The correlation matrix indicates few notably high correlations that are rather intuitive. The highest correlations displayed are between volatility and MAX, and fund TNA and company TNA. This is expected since a large MAX often leads to higher volatility. The other high correlation indicates the natural relationship of large fund companies having larger funds.

The correlation between volatility and MAX could be an issue in the regression since a high collinearity can affect the standard errors. However, AG (2020) alleviate potential concerns that this would affect the results by testing two alternative measures of MAX, a matched-adjusted MAX[16] and a residual

[13] The front load fee is an upfront percent charge or an upfront fixed charge on allocated capital. The rear load fee is a percent charge or an upfront fixed charge on withdrawn capital.

[14] In the CRSP dataset, it is possible to see the load fees based on different levels of capital allocation per customer, but it is impossible to identify what load fees that have actually been applied, since we do not know the distribution of capital allocation.

[15] AG (2020) dataset contains 150,181 fund quarter observations from 3,674 distinct funds. Note that their dataset contains ~ 6 more years. Some differences may also stem from differences in data filtering. In a perfect replication setting, we would have used the same dataset as AG (2020), but we have been unable to obtain that dataset.

[16] Using benchmark portfolios in each quarter, where funds are ranked based on volatility, the

MAX[17]. These alternative measures of MAX still allow a positive and significant relation to fund flows, suggesting that collinearity is an unlikely driver of the results.

A Pearson Chi-squared test on the MAX and MIN frequencies indicates that the data displays seasonality in the distribution of MAX and MIN over months[18]. We can reject the null hypothesis (both for MIN and MAX) that the distributions would be uniform at a significance level of 0.1% (chi-square test statistics are both >4,000). Seasonality could arise due to several factors, one of them being the phenomenon of "window dressing" at year- or quarter-ends. The idea behind window dressing is that fund performance is not only dependent on actual performance but also evaluated based on screenings of holdings at year-end or quarter-end (Lakonishok et. al. 1991). A fund manager can therefore increase their share of winner stocks or decrease their share of losing stocks to mislead the evaluator into thinking that returns must be good, since they seem to have a portfolio with a high share of stocks that have been "winners" and low share of stocks that have been "losers".

Our distribution over months differs 6% for MIN and 5% for MAX between the most frequent month and the most infrequent month. This displays a slightly less smooth distribution over months relative to AG (2020)19.

matched MAX is the difference between a fund's MAX and the average MAX in the benchmark portfolio that a fund belongs to that quarter

[17] Replacing the MAX with the residuals from a cross-sectional regression of MAX trailing twelve month returns and volatility.

[18] The Pearson Chi-squared test statistic is 4,600 for MIN and 4,000 for MAX. We can therefore reject the null hypothesis that the MAX and MIN frequencies (respectively of course) occur equally likely over different months.

[19] Our distribution over months differ 6% for MIN and 5% for MAX, whereas the difference is almost 4% for AG (2020).

Table 2. Mean values[20] of variables over different years.

Descriptive statistics - mean by year

	FLOW	AGE	VOL	SKEW	FUND_ TNA	COMP _TNA	MGMT _FEE	EXP_ RATIO	TURN_ RATIO	MIN	MAX
1999	.209	33.878	.05	.239	740.3	18008.8	.007	.015	1.088	-.046	.043
2000	.074	95.462	.068	.396	1662.9	28399.4	.007	.015	.857	-.07	.071
2001	.001	99.221	.068	.157	1259.4	22008.6	.007	.015	.885	-.054	.055
2002	-.013	97.05	.059	-.059	1127.5	20019.7	.007	.016	.883	-.039	.036
2003	.061	99.072	.051	-.011	1505.3	23540.0	.007	.015	.794	-.03	.031
2004	.053	102.427	.032	-.072	2087.0	31579.3	.007	.015	.776	-.021	.022
2005	.029	106.022	.034	-.09	2271.5	36096.6	.007	.015	.703	-.021	.022
2006	.023	115.151	.03	-.216	2677.1	42196.6	.007	-.007	.705	-.02	.021
2007	.021	117.43	.027	-.523	2904.5	49436.8	.007	-.001	.836	-.018	.019
2008	-.035	118.375	.052	-.359	2286.8	38929.6	.007	-.044	.921	-.028	.027
2009	.055	124.691	.083	-.414	1950.1	34533.4	.007	-.016	.711	-.034	.035
2010	.05	128.43	.055	-.265	2367.2	46173.8	.007	-.011	.668	-.021	.022
2011	-.012	133.458	.05	-.058	2529.8	51476.6	.007	.013	.565	-.019	.02
2012	.026	138.492	.05	-.059	2583.7	54306.3	.007	.008	.534	-.02	.021
2013	.049	142.678	.028	-.383	3017.1	66645.6	.007	-.057	.549	-.016	.017
2014	.014	143.697	.031	-.383	3362.7	78065.5	.007	.012	.573	-.019	.018
2015	-.012	149.362	.034	.179	3373.3	77008.6	.006	-.103	.62	-.019	.019
2016	.01	155.062	.041	.099	3304.8	73352.0	.006	-.01	.583	-.021	.022

Table 3. Correlation matrix

Matrix of correlations

Variables	FLOW	LN_ AGE	VOL	SKEW	FUND_ TNA	COMP_ TNA	LAG_ FLOW	MGMT _FEE	EXP_ RATIO	TURN_ RATIO	MIN	MAX	LOW_ PERF	HIGH_ PERF
FLOW	1.000													
LN_AGE	-0.235	1.000												
VOL	0.012	-0.038	1.000											
SKEW	-0.000	-0.026	0.120	1.000										
FUND_TNA	-0.067	0.238	-0.080	-0.005	1.000									
COMP_TNA	-0.006	0.066	-0.080	-0.041	0.619	1.000								
LAG_FLOW	0.409	-0.259	0.012	0.001	-0.034	0.007	1.000							
MGMT_FEE	-0.013	0.023	0.117	0.022	-0.179	-0.345	-0.009	1.000						
EXP_RATIO	-0.064	-0.051	0.107	0.016	-0.231	-0.224	-0.056	0.456	1.000					
TURN_RATIO	-0.002	-0.063	0.090	0.010	-0.167	-0.112	0.004	0.160	0.150	1.000				
MIN	0.023	0.114	-0.368	-0.072	0.166	0.191	0.015	-0.195	-0.200	-0.195	1.000			
MAX	0.027	-0.032	0.787	0.123	-0.046	-0.060	0.031	0.074	0.053	0.052	-0.182	1.000		
LOW_PERF	-0.080	-0.001	0.021	0.007	-0.027	-0.046	-0.005	0.055	0.062	0.031	-0.093	0.020	1.000	
HIGH_PERF	0.088	0.004	0.016	-0.004	-0.042	-0.035	0.013	0.058	0.016	0.027	-0.065	0.028	-0.253	1.000

[20] Note that the mean values are presented as a simple average of all data in the dataset. The variables in this table are therefore not weighted with total net assets and can therefore not be compared directly with AG (2020).

Figure 3. Distribution of what month that the MIN for each data point corresponds to. For example, 12% of data points have a MIN that occurred in January.

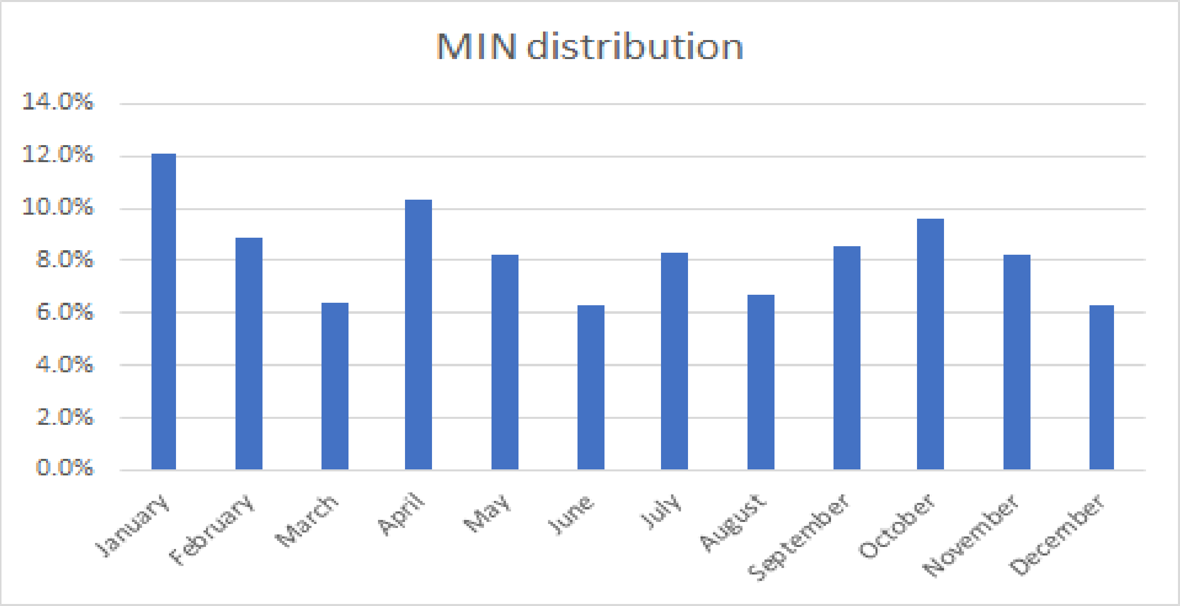
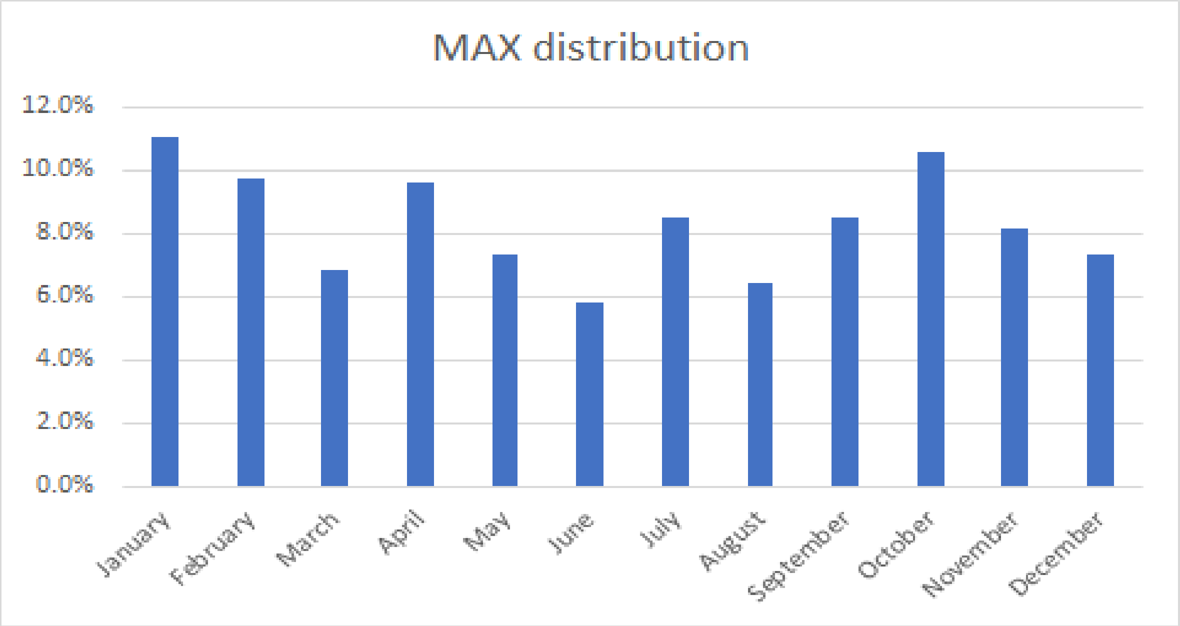


Figure 4. Distribution of what month that the MAX for each data point corresponds to.



5. The regression model

Like AG (2020), we employ the Fama and Macbeth (1973) regression model to test the relationship between MIN and FLOW in the subsequent quarter. The Fama and Macbeth regression (1973) is a two-step regression model that is suitable for handling time series panel data. The first step is to regress the dependent variable (in our case FLOW) against all proposed control factors. This produces a coefficient for each factor for each share class. Step two is to regress FLOW against these produced coefficients to achieve an overall coefficient for each control factor. The model was originally developed to estimate asset prices/returns/excess returns based on risk factors. In all regressions that we run, all independent variables except for AGE are lagged with one quarter, to better represent what information that investors have available.

$$FLOW_{i,t} = \alpha + \beta_{i,t-3} * MIN_{i,t-3} + \delta_{i,t-3} * X_{i,t-3} + \varepsilon_{i,t-3}$$

Where the second control variable term X is a vector of the following control variables:

$$\begin{gathered} \ln\left(FUND\_TNA_{i,t-3}\right) \\ \ln\left(COMP\_TNA_{i,t-3}\right) \\ \ln\left(AGE_{i,t}\right) \\ MGMT\_FEE_{i,t-3} \\ TURN\_RATIO_{i,t-3} \\ EXP\_RATIO_{i,t-3} \\ VOL_{i,t-3} \\ SKEW_{i,t-3} \\ FLOW_{i,t-3} \\ LOW\_PERF_{i,t-3} \\ HIGH\_PERF_{i,t-3} \end{gathered}$$

Note that time is still denoted in months, meaning that time index $t-3$ indicates the ending month of the previous quarter. The reader is reminded that the variables still reflect quarterly data. For example: $FLOW_{i,t}$ denotes flow for the entire previous quarter with ending month t. $FLOW_{i,t-3}$ denotes flow for the entire quarter with ending month $t-3$. To avoid perfect multicollinearity, MID_PERF is the baseline category.

We employ the following regression model to test the relationship between MAX and FLOW:

$$FLOW_{i,t} = \alpha + \beta_{i,t-3} * MAX_{i,t-3} + \delta_{i,t-3} * X_{i,t-3} + \varepsilon_{i,t-3}$$

We employ the following regression model to test the relationships between MIN and FLOW and MAX and FLOW simultaneously:

$$FLOW_{i,t} = \alpha + \beta_{i,t-3} * MIN_{i,t-3} + \gamma_{i,t-3} * MAX_{i,t-3} + \delta_{i,t-3} * X_{i,t-3} + \varepsilon_{i,t-3}$$

6. Results

6.1 MIN - FLOW Relationship

The regression results indicate that there is a positive relationship between increasing (less extreme) MIN and FLOW. This implies that investors tend to allocate more capital to funds with less extreme historically negative returns. Without MAX included in the regression (1), a +1 percentage point change in MIN yields an increase of 0.42 percentage points in FLOW. This relationship is robust to adjustment for fund performance and is significant at a 0.1% level[21]. In the combined regression (3), where both MIN and MAX are included as independent variables, the effect of MIN on fund flow is still significant at the 0.1% level. With MAX included in the regression, a +1 percentage point change in MIN yields an increase of 0.71 percentage points in FLOW. Hence, compared to regression 1 (without MAX), the size of the MIN coefficient is larger in the combined setting.

In order to compare the effects between control variables, we use standardized control variables in the last three regressions, in accordance with AG (2020). By doing this, the coefficient estimate of a standardized control variable responds to the effect on FLOW that a change in the independent variable equal to one (1) standard deviation has. For reference, the average FLOW in a quarter is 2.3%. For the MIN-FLOW relationship, an increase by one (1) standard deviation[22] yields a 1.1 percentage points increased FLOW in the next quarter.

In the standardized variables combined regression (6), we can compare the effect of MIN and MAX, relative to their respective deviations in the dataset. We notice that MAX, considering its larger deviation in our dataset, seems to have a larger effect (0.066) on FLOW than MIN has (0.018). This is explained by the fact that although the coefficients of MIN and MAX are somewhat similar in regressions (1) and (2), MAX is more volatile (as we notice by observing their standard deviations, MIN: 2.1% and MAX: 8.7%).

[21] Note that Table 2 shows *** to denote the 1% level, but t-values are included

[22] For reference, the standard deviation of MIN is 2.1%.

Table 4. Regression results for the three non-standardized regressions and the three standardized regressions.

Regression results

	(1) MIN	(2) MAX	(3) Combined	(4) Standard. MIN	(5) Standard. MAX	(6) Standard. Combined
MIN	0.416***		0.707***	0.011***		0.018***
	(6.102)		(6.721)	(6.102)		(6.721)
MAX		0.528***	0.820***		0.043***	0.066***
		(7.995)	(7.402)		(7.995)	(7.402)
VOL	0.066	-0.547***	-0.282**	0.002	-0.017***	-0.009**
	(0.491)	(-4.552)	(-2.231)	(0.491)	(-4.552)	(-2.231)
SKEW	0.001	-0.000	-0.005	0.001	-0.000	-0.002
	(0.275)	(-0.075)	(-1.344)	(0.275)	(-0.075)	(-1.344)
LN_AGE	-0.030***	-0.030***	-0.030***	-0.023***	-0.022***	-0.023***
	(-23.490)	(-23.929)	(-23.750)	(-23.490)	(-23.929)	(-23.750)
LN_FUND_TNA	-0.004***	-0.004***	-0.004***	-0.008***	-0.007***	-0.007***
	(-10.714)	(-10.887)	(-10.663)	(-10.714)	(-10.887)	(-10.663)
LN_COMP_TNA	0.002***	0.003***	0.002***	0.004***	0.006***	0.005***
	(5.870)	(7.176)	(6.528)	(5.870)	(7.176)	(6.528)
EXP_RATIO	-1.529***	-1.602***	-1.499***	-0.009***	-0.010***	-0.009***
	(-8.212)	(-8.451)	(-7.773)	(-8.212)	(-8.451)	(-7.773)
TURN_RATIO	-0.000	-0.001	-0.000	-0.000	-0.001	-0.000
	(-0.148)	(-1.017)	(-0.278)	(-0.148)	(-1.017)	(-0.278)
MGMT_FEE	1.477***	1.144***	1.151***	0.004***	0.003***	0.003***
	(7.097)	(5.588)	(5.503)	(7.097)	(5.588)	(5.503)
LAG_FLOW	0.343***	0.341***	0.336***	0.064***	0.064***	0.063***
	(33.962)	(34.562)	(33.793)	(33.962)	(34.562)	(33.793)
LOW_PERF	-0.021***	-0.025***	-0.022***	-0.008***	-0.010***	-0.009***
	(-8.590)	(-10.786)	(-9.722)	(-8.590)	(-10.786)	(-9.722)
HIGH_PERF	0.031***	0.027***	0.028***	0.012***	0.011***	0.011***
	(9.236)	(8.922)	(9.000)	(9.236)	(8.922)	(9.000)
Obs.	330307	330307	330307	330307	330307	330307
R-squared	0.238	0.239	0.243	0.238	0.239	0.243

T-values are in parenthesis

**** p<0.01, ** p<0.05, * p<0.1*

6.2 MAX-FLOW Relationship

The main research contribution of our article is the MIN-FLOW relationship. However, in order to compare our results with those of AG (2020), we argue that the MAX-FLOW relationship is interesting to present as well. The time-strained reader who is only interested in the MIN-FLOW relationship can therefore skip this section.

The MAX-FLOW relationship is robust to adjustment for fund performance and is also significant at a 0.1% level. Regression 2 shows that without MIN included in the regression, a +1 percentage point change in MAX yields an increase of 0.53 percentage points in FLOW. We can compare this to the effect between 0.26 to 0.33 in AG's (2020) regressions, which indicates that our regression results are in fact quite similar, although our regression indicates a slightly stronger relationship between MAX and FLOW. In the standardized variable regression (5), an increase in MAX by 1 standard deviation[23] yields +4.3 percentage points increased FLOW in the next quarter. To compare with AG (2020), their result is that 1 standard deviation yields 0.75% (piecewise linear regression[24]) and 0.95% (quadratic regression[25]). However, the standard deviation of MAX most likely differs between our dataset and AG (2020), meaning that it might be more appropriate to compare the regression without standardized variables. AG (2020) do not present the standard deviation, which prohibits us from comparing our results with theirs in the standardized variable regression.

In the combined regression (3), a +1 percentage point change in MAX yields an increase of 0.82 percentage points in FLOW.

6.3 Robustness of the regression

The Fama and Macbeth (1973) regression does not have a conventional Variance Inflation Factor test. Instead, we rely on the correlation matrix to detect possibly critical multicollinearity, in order to test the quality of our model. We also run a Variance Inflation Factor test on a cross-sectional regression, instead of the Fama and Macbeth (1973), with the same data and variables as the above regression. This test indicates that there is no critical multicollinearity in either the MIN-FLOW or the MAX-FLOW regressions, as VIF is below 3.5 for all variables in the model.

Most control variables show expected signs across the regressions. For a more detailed evaluation, see table 5 in Appendix II.

Since our model is based on share class data of funds, it is necessary to use robust standard errors. However, since we also want to adjust

[23] For reference, the standard deviation of MAX is 8.7%.
[24] Piecewise linear regression refers to ranking performance as LOW, MIDDLE and HIGH.
[25] Quadratic regression refers to ranking performance with one continuous linear and one quadratic term.

standard errors for autocorrelation in the time series data, we employ the Newey and West (1987) standard errors with 1 time-period lag. The Newey and West standard error estimation is especially used to adjust standard errors for autocorrelation, but is also robust to heteroscedasticity in standard errors (Newey and West, 1987), and performs very similar to Huber-White robust standard error estimation for panel data (Petersen, 2009). The Newey and West (1987) standard errors are also used by AG (2020).

AG (2020) perform a separate test to check whether their results are robust when using clustered standard errors. We have not been able to test the robustness of our results with clustered standard errors as we have used the "asreg" package in STATA to perform the Fama and Macbeth (1973) regression. This package does not allow Newey and West (1987) standard errors to be clustered. Although, AG (2020) conclude that their results are indeed robust when using clustered standard errors, it should be mentioned that since our data is on a share class level, it would be even more interesting for us to perform this test.

7. Discussion and Implications

7.1 Interactions between MIN and MAX

Compared to the original model developed by AG (2020), we introduce an additional variable of interest into the regression, MIN. From a theoretical point of view, MAX and MIN should both impact fund flow, as they are tail events on different sides of the loss and gain domain. According to cumulative prospect theory, extreme MINs should lead to lower fund flow, while extreme MAXs should lead to higher fund flow. However, intuitively one might think that a fund with a high MAX could also have a high (less extreme) MIN, implying that the effect we see from an extreme MIN could simply be attributed to MIN functioning as a proxy variable for MAX. Therefore, we need to ensure that the effect of MIN on fund flow displayed in the regression is not simply a consequence of correlation with MAX. We check this is in two ways. First, by examining the combined regression it becomes evident that both MIN and MAX are significant also in a combined setting. Second, a negative correlation can be observed between MIN and MAX in the correlation matrix. Since the correlation is negative, we would expect MIN and MAX to show coefficients with different signs, if most of the effect of MIN would be that it functions as a proxy for MAX. As the reader can observe in the regression, the coefficients have the same sign. To conclude, we seem to find rather strong statistical evidence that the effect of an extreme MIN is a real and standalone effect, and not just a proxy for the effect of an extreme MAX.

7.2 Theoretical implications: Prospect theory as an explanatory model of our results

It is perhaps counterintuitive that investors in mutual funds would exhibit gambling-like tendencies by preferring funds with lottery-like returns, as an important objective for these investors would be to benefit from diversification and liquidity. However, AG (2020) suggest that investors' tendencies to overweight positive extreme returns is a more widespread phenomenon throughout financial markets than what has previously been understood, as it is also present for individuals who have *already* decided to invest in mutual funds. Building on the work by AG (2020), our research suggests that investors also disfavor funds with the opposite distributional feature, namely extreme negative returns on record. Our results, together with the result from AG (2020), therefore suggest that mutual fund investors are subject to erring investment decisions, at least from an expected utility perspective, on both the likelihood of high-payoff states as well as low-payoff states.

Our results are in line with cumulative prospect theory, which predicts that individuals overweight tail-events in their decision-making. In our case, the theory predicts that investors would be less inclined to buy funds with more extreme MINs, which is also what our empirical results show. Our results, indicating that investors prefer funds with less extreme MINs, might however not be as counterintuitive as the results that investors prefer funds with extreme MAXs. A key reason to invest in mutual funds would be to lower the risk by benefitting from diversification, and the investors of mutual funds could therefore be expected to be risk averse, and hence react more strongly to the prospect of heavy losses. However, as we control for volatility, the effect of MIN goes beyond the "rational" (or at least commonly quantified) way in which investors shy away from risk. We can therefore argue that investors that would want to control their risk levels would most likely do so by observing the volatility of the fund, as opposed to the MIN of the fund. As a result, trading on MIN should not be considered a rational investment strategy, even for loss averse investors.

An interesting result from our regression is that the coefficient for MAX is larger than the coefficient for MIN, implying that the effect on fund flow is stronger for a change in MAX compared to MIN. The standardized variable combined regression (6) suggests that the overall impact of MAX on fund flow seems larger than the overall impact of MIN on fund flow, when accounting for their deviations. These results imply that MAX is more important than MIN to explain future fund flows.

As the coefficient of MAX is larger than the coefficient of MIN in our regression, our results stand in contrast to the results of Kahneman and Tversky (1979, 1992), where individuals are shown to exhibit substantial loss aversion. In fact, most studies find loss aversion coefficients[26] around 2, meaning losses weigh approximately twice as much as gains (Tversky and Kahneman, 1992; Abdellaoui, Bleichrodt and Kammoun, 2013; Abdellaoui *et al.*, 2016). In contrast to this, our results indicate that the gambling-behavior associated with a preference for an extreme MAX (AG, 2020) is stronger compared to the loss aversion associated with a preference for a non-extreme MIN. Interestingly, other researchers have found the same phenomenon when prospect theory is studied in a non-laboratory environment, implying that the utility function differs from the predictions of classic prospect theory (Kahneman and Tversky, 1979). As an example, Abdellaoui et. al. (2013) encountered surprising results when trying to replicate laboratory results with observations of financial professionals. Specifically, they found that a large portion of the financial professionals in their study showed gain seeking behavior (as opposed to risk aversion). On aggregate, their study also showed much lower levels of loss aversion compared to other studies.

There is however a possible explanation to our results concerning the relative effect of MIN and MAX on fund flow, that would not violate the fundamentals of cumulative prospect theory. Recall that fund flow is not only driven by investors evaluating whether to enter the fund or not, but also by investors already invested in the fund, evaluating whether to decrease or increase their fund holdings. As it turns out, while “outside” investors would only be subject to probability weighting within cumulative prospect theory (overweighting extreme events), “inside” investors would experience an additional effect depending on whether they are in the loss or gain domain. According to the S-shaped value function in cumulative prospect theory (Kahneman and Tversky, 1992), the risk-aversion is context-dependent for inside investors, who will be risk-seeking in the loss domain and risk-averse in the gain domain. For inside investors, an extreme MIN likely causes the investor to fall into the loss domain[27] whereas an extreme MAX likely causes the investor to fall into the gain domain. This implies that inside investors would be risk-seeking in relation to an extreme MIN, meaning they might not contribute to a negative fund flow by selling their position. Respectively, inside investors would be risk-averse in relation to an extreme MAX, meaning they would sell their position in the gain domain, contributing to a negative fund flow. Hence, inside investors will likely have a directly opposite response compared to outside investors. Since the effect is larger in the loss

[26] Note that the loss aversion coefficient is not directly comparable to our results, we do not calculate this number in our statistical analysis

[27] Note that being in the loss domain does not necessarily require that the investor would have a negative total accumulated return, instead the perceptions of loss could also be valid for losing a substantial part of a previously accumulated gain, such as with an extreme MIN.

domain (the S-curve is steeper) this could contribute to explaining why the effect of MAX is larger than the effect of MIN on fund flow. Unfortunately, as stated, the data does not allow us to distinguish between inside and outside investors, nor between inside investors in loss domain or gain domain. Therefore, we cannot expand this theoretical reasoning into an empirical test.

7.3 Accessibility to MIN and MAX in investors' option space

In order for prospect theory to be a plausible explanatory model for the observed effects of extreme MAX and MIN on fund flow, investors must be able to consider MAX and MIN in their decision-making. This is only possible if information regarding MAX and MIN are accessible to investors. AG (2020) argue that the adoption of online platforms to access financial data has made information about MAX available to investors, as many financial sites report historical lowest and highest monthly returns from the recent year. If this is the case, the relationship between fund flow and MAX (and MIN) could be subject to a time-component and this relationship should grow stronger with time, as more individual investors can access financial information through online services. If the relationship does not grow stronger with time, AG's (2020) explanation of the relationship between MAX and fund flow, which is based on cumulative prospect theory, could be questioned. In this case, investors would have to access MAX in some other way, or it would be difficult to argue that cumulative prospect theory would be the correct explanatory model. The effect of time on the relationship between MAX and fund flow and MIN and fund flow would be interesting to test. However, the regression model specifications of how to test this would not be trivial, the main problem being that the interactions between time and MAX, as well as time and MIN, could consist of a wide range of other effects apart from the effect that we actually would like to measure, the growth of online investment research. Therefore, any time-series over this data period could be tainted by other external factors that affected the financial markets over these years. Examples of such external factors could be the dot com bubble and the 2008 financial crisis that have affected investment behavior and capital allocation over this time period. Reliably measuring the effect of investors' access to information about MIN and MAX is something that we therefore leave for future research.

Another interesting consideration regarding investors' access to information about MAX and MIN, is our use of style-adjusted return and not the actual return (in this paper as well as the research done by AG (2020)). In line with the accessibility prerequisite, investors must have access to information about MAX and MIN in order to include it in their decision making. There is no doubt that investors have easier access to actual return, rather than style-adjusted return, as style-adjusted return is seldomly directly reported. This raises the question whether investors do in fact have access to MAX and MIN if they cannot observe the style-adjusted return. However, when doing investment research, investors are often presented with securities, such as mutual funds, that are grouped by categories, for example; geographical market, industry focus or investment style. This grouping of information means that although investors are not directly exposed to style-adjusted returns, they are often presented with the actual return of a fund as well as the actual returns of comparable funds. Hence, the evaluation of fund return is often subject to benchmarking. Furthermore, previous research suggests that investment decisions are likely to be based on fund style (Mullainathan, 2002; Barberis and Shleifer, 2003; Pomorski, 2011). This would imply that investors compare the return of funds within the same fund style when investing (rather than comparing the return between funds in general) if they have decided on the style before deciding on the specific fund. We therefore argue that investors do have access to style-adjusted return and that both MAX and MIN fulfill the accessibility prerequisite.

7.4 Practical implications: Consequences of the MIN-FLOW relationship on portfolio strategy

In order to understand why our findings are relevant, one needs to understand why the capital flow in and out of a fund (fund flow) is important. From previous literature, we can identify two main reasons expanded on below:

7.4.1 MAX as a portfolio strategy?

First, the fund manager's incentives are usually based on assets under management (AUM), implying that fund flow is a driver of fund manager behavior. In line with this, prior research shows that fund flow impacts the fund manager's behavior. For example, as return is a determining factor for future fund flow, fund managers tend to alter the riskiness of their portfolios as a response to the relationship between performance and fund flow, especially towards the end of the year (Chevalier and Ellison, 1997; Brown, Harlow and Starks, 1996). However, fund flow is not always strictly related to previous fund performance. Research by Cooper, Gulen and Rau (2005) implies that if a fund changes name to something in line with a "hot and new" investment style, they can increase fund flow. Financial ratings, media attention, in-your-face fees and marketing efforts have also been shown to be determinants of fund flow, adding to the idea that the fund flow is not always linked to measures such as fund performance (Sirri and Tufano, 1998; Jain and Shuang Wu, 2000; Barber, Odean and Zheng, 2005; Guercio and Tkac, 2008; Gallaher, Kaniel and Starks, 2009). Hence, cosmetic effects can influence capital allocation in funds for consumers in an irrational way. Fund managers can therefore also employ cosmetic measures to manipulate fund flow.

As our results suggest that MIN and MAX impact fund flow, this raises the question of how fund flow could be manipulated by considering MIN and MAX in the portfolio strategy. In fact, in the conclusion by AG (2020), they introduce the idea that fund managers could increase their fund flow by choosing a portfolio strategy that would generate more extreme MAXs. As fund managers can go to great lengths to manipulate fund flow (Chevalier and Ellison, 1997; Brown, Harlow and Starks, 1996), the idea of a MAX-based portfolio strategy becomes highly interesting. The results we show in this thesis can contribute to an added perspective on such a strategy. Specifically, our results shed light on the fact that fund managers who seek to generate extreme MAXs would also have to take into consideration how the effect of an extreme MIN would affect fund flow. As MAX and MIN show a negative correlation of -0.2, a strategy in place to generate extreme MAXs would likely also generate a higher relative portion of extreme MINs. Since investors, according to our regression, react negatively to extreme MINs, a fund management strategy to generate extreme MAXs in order to achieve higher fund flow, would at least partially be countered by the negative effect on fund flow from the effect of extreme MINs. From this, our research contributes to the interesting discussion introduced by AG (2020) regarding portfolio management considerations.

7.4.2 Ability to predict flow-induced trades

Second, fund flow also plays a key role in the tendency for mutual fund returns to persist over multi-year periods, meaning ‘winning’ funds have a larger tendency to keep on being winning funds, while “losing” funds have a larger tendency to keep on being losing funds (Wermers, 2003). Although part of the momentum can be ascribed to the fact that winning funds hold winning stocks, the consumer reaction to fund performance and the resulting behavior of the fund managers both play important roles. While managers of losing funds tend to be reluctant to sell their low return stocks in favor of momentum stocks, the higher cash inflows of the winning funds enable these to further implement momentum strategies (Wermers, 2003; Grinblatt and Han, 2005).

This also has implications for underlying stock prices. As the empirical research suggests that fund flow affects underlying stock prices as winning and losing funds respectively often have correlated holdings due to portfolio limitations imposed by the fund style. As funds with sizable capital inflows (winning funds) will be inclined to expand their current positions, the stocks held in common by these funds will be subject to a positive price pressure. Likewise, funds with sizable outflows (losing funds) will be inclined to decrease their existing positions. Hence, stocks held in common by constrained funds will be subject to downward price pressure (Wermers, 2003;

Coval and Stafford, 2007). Coval and Stafford (2007) therefore suggest that a trading strategy based on trading against distressed mutual funds can yield significant returns. As a result, there is an incentive to front-run[28] the expected flow-induced trades by constrained funds, making predictions of future fund flows highly relevant. Hence, taking MIN (and MAX) into account could enhance the performance of any financial model used to predict future fund flow.

8. Conclusion

Our results show that investors in actively managed equity funds make their investment decisions in line with one of the predictions of cumulative prospect theory, namely that investors place excess weight on tail events. This implies that negative extreme payoff states in the historical return will result in smaller capital inflow in the future, even when the results are adjusted for historical performance, volatility, fund fees, fund size and company size. Interestingly, our findings indicate that MAX has a larger impact than MIN on future fund flow. Our results have implications both from a theoretical and practical point of view. First, we contribute to the line of research investigating how the predictions of cumulative prospect theory perform in a real-life setting. Second, we contribute to a deepened understanding of the mechanisms affecting fund flow, with implications for two trading strategies.

Regarding our theoretical contribution, although we observe a relationship between MIN and fund flow, in line with the predictions of cumulative prospect theory, we cannot prove that this relationship is actually a result of said theory. Hence, our discussion regarding the mechanisms behind the observed effect on fund flow due to extreme negative payoff states is a theoretical one. In order to strengthen the argument that cumulative prospect theory is the correct explanatory model to use when accounting for the mechanisms behind the observed effects, we suggest two ideas for future research.

First, we suggest that future research should try to answer if MIN and MAX are actually accessible in investors' option space, for instance by constructing a dataset where the magnitude of the effect over time can be studied in relation to the adoption of online financial tools. Second, our inability to distinguish "inside" investors from "outside" investors in the dataset restrains our ability to relate our results to prospect theory. If investor behavior can be studied in the gain and loss domain respectively, the predictions of prospect theory could be studied with more accuracy.

Regarding our practical contribution, the consequences of the two discussed trading

[28] In this context, front-running is not used to describe the illegal act of trading on advance information. Instead, the term is used to describe the procedure of trying to predict stock price movements by anticipating the behavior of large mutual funds.

strategies should be studied in more depth. Furthermore, in order to protect retail investors from financial deterioration, more research is needed to understand how such investors could be safeguarded from erring investment decisions based on extreme MINs and MAXs.

9. References

Abdellaoui, M. *et al.* (2016) 'Measuring Loss Aversion under Ambiguity: A Method to Make Prospect Theory Completely Observable', *Journal of Risk and Uncertainty*. Springer New York LLC, 52(1). doi: 10.1007/s11166-016-9234-y.

Abdellaoui, M., Bleichrodt, H. and Kammoun, H. (2013) 'Do financial professionals behave according to prospect theory? An experimental study', *Theory and Decision*, 74, pp. 411–429. doi: 10.1007/s11238-011-9282-3.

Akbas, F. and Genc, E. (2020) 'Do Mutual Fund Investors Overweight the Probability of Extreme Payoffs in the Return Distribution?', *Journal of Financial and Quantitative Analysis*. Elsevier BV, 55(1), pp. 223–261. doi: 10.2139/ssrn.2734743.

Barber, B. M. and Odean, T. (2012) 'The Behavior of Individual Investors', *SSRN Electronic Journal*. Elsevier BV. doi: 10.2139/ssrn.1872211.

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), pp. 2095–2119. doi: 10.1086/497042.

Barberis, N. (2013) 'The psychology of tail events: Progress and challenges', *American Economic Review*, 103(3), pp. 611–616. doi: 10.1257/aer.103.3.611.

Barberis, N. C. (2013) 'Thirty Years of Prospect Theory in Economics: A Review and Assessment', *Journal of Economic Perspectives*, 27(1). doi: 10.1257/jep.27.1.173.

Barberis, N. and Shleifer, A. (2003) 'Style Investing', *Journal of Financial Economics*, 68(2).

Barberis, N. and Huang, M. (2008) 'Stocks as lotteries: The implications of probability weighting for security prices', *American Economic Review*, 98(5), pp. 2066–2100. doi: 10.1257/aer.98.5.2066.

Brown, K. C., Harlow, W. V. and Starks, L. T. (1996) 'Of Tournaments and Temptations: An Analysis of Managerial Incentives in the Mutual Fund Industry', *The Journal of Finance*. JSTOR, 51(1), p. 85. doi: 10.2307/2329303.

Brunnermeier, M. K., Gollier, C. and Parker, J. A. (2007) 'Optimal beliefs, asset prices, and the preference for skewed returns', *American Economic Review*, 97(2), pp. 159–165. doi: 10.1257/aer.97.2.159.

Chevalier, J. and Ellison, G. (1997) 'Risk taking by Mutual Fund as a response to incentives', *Journal of Political Economy*. University of Chicago Press, 105(6), pp. 1167–1200. doi: 10.1086/516389.

Cooper, M. J., Gulen, H. and Rau, P. R. (2005) 'Changing Names with Style: Mutual Fund Name Changes and Their Effects on Fund Flows', *The Journal of Finance*, 60(6), pp. 2825–2858. doi: 10.2307/3694805.

Coval, J. and Stafford, E. (2007) 'Asset fire sales (and purchases) in equity markets', *Journal of Financial Economics*. North-Holland, 86(2), pp. 479–512. doi: 10.1016/j.jfineco.2006.09.007.

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*, LVI (6).

Eraker, B. and Ready, M. (2015) 'Do investors overpay for stocks with lottery-like payoffs? An examination of the returns of OTC stocks', *Journal of Financial Economics*. Elsevier, 115(3), pp. 486–504. doi: 10.1016/j.jfineco.2014.11.002.

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

Fama, E., & MacBeth, J. (1973). 'Risk, Return, and Equilibrium: Empirical Tests'. *Journal of Political Economy*, 81(3), pp. 607-636.

Fred van Raaij, W. (2016) 'Understanding Consumer Financial Behavior: Money Management in an Age of Financial Illiteracy'. New York City: Palgrave Macmillan. doi: 10.1057/9781137544254.

De Giorgi, E. G. and Legg, S. (2012) 'Dynamic portfolio choice and asset pricing with narrow framing and probability weighting', *Journal of Economic Dynamics and Control*. Elsevier, 36(7), pp. 951–972. doi: 10.1016/j.jedc.2012.01.010.

Gallaher, S. T., Kaniel, R. and Starks, L. T. (2009) 'Advertising and Mutual Fund: From Families to Individual Funds', *SSRN Electronic Journal*. Elsevier BV. doi: 10.2139/ssrn.1362535.

Grinblatt, M. and Han, B. (2005) 'Prospect theory, mental accounting, and momentum', *Journal of Financial Economics*, 78, pp. 311–339. doi: 10.1016/j.jfineco.2004.10.006.

Guercio, D. Del and Tkac, P. A. (2008) 'Star power: The effect of morningstar ratings on Mutual Fund flow', *Journal of Financial and Quantitative Analysis*. Cambridge University Press, 43(4), pp. 907–936. doi: 10.1017/S0022109000014393.

Investment Company Institute (2019) *FACT BOOK A Review of Trends and Activities in the Investment Company Industry 59th edition*.

Jain, P. C. and Shuang Wu, J. (2000) 'Truth in Mutual Fund Advertising: Evidence on Future Performance and Fund Flows', *The Journal of Finance*, 55(2), pp. 937–958.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analsysis of Decision Under Risk. Econometrica (pre-1986), 47(2), 263–291.

Lakonishok, Josef, Andrei Shleifer, Richard H Thaler, and Robert W Vishny. 1991. "Window Dressing by Pension Fund Managers." American Economic Review Papers and Proceedings May.

Mullainathan, S. (2002) *Thinking Through Categories*. MIT and NBER.

Newey, W., & West, K. (1987). 'A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix'. *Econometrica*, 55(3), 703-708. doi:10.2307/1913610

O'Brien, R. M. (2007) 'A caution regarding rules of thumb for variance inflation factors', *Quality and Quantity*, 41(5), pp. 673–690. doi: 10.1007/s11135-006-9018-6.

Pomorski, L. (2011) 'Style Investing: Evidence from Mutual Fund Flows', *SSRN Electronic Journal*. Elsevier BV. doi: 10.2139/ssrn.559405.

Sensoy, B. A. (2008) 'Performance evaluation and self-designated benchmark indexes in the Mutual Fund industry $'. doi: 10.1016/j.jfineco.2008.02.011.

Sirri, E. R. and Tufano, P. (1998) 'Costly Search and Mutual Fund Flows', *The Journal of Finance*, LIII (5).

Stott, H. P. (2006) 'Cumulative prospect theory's functional menagerie', *Journal of Risk and Uncertainty*. Springer, 32(2), pp. 101–130. doi: 10.1007/s11166-006-8289-6.

Tversky, A. and Fox, C. R. (1995) 'Weighing risk and uncertainty', *Psychological Review*. American Psychological Association Inc., 102(2), pp. 269–283. doi: 10.1037/0033-295X.102.2.269.

Tversky, A. and Kahneman, D. (1981) 'The Framing of decisions and the psychology of choice', *Science*.

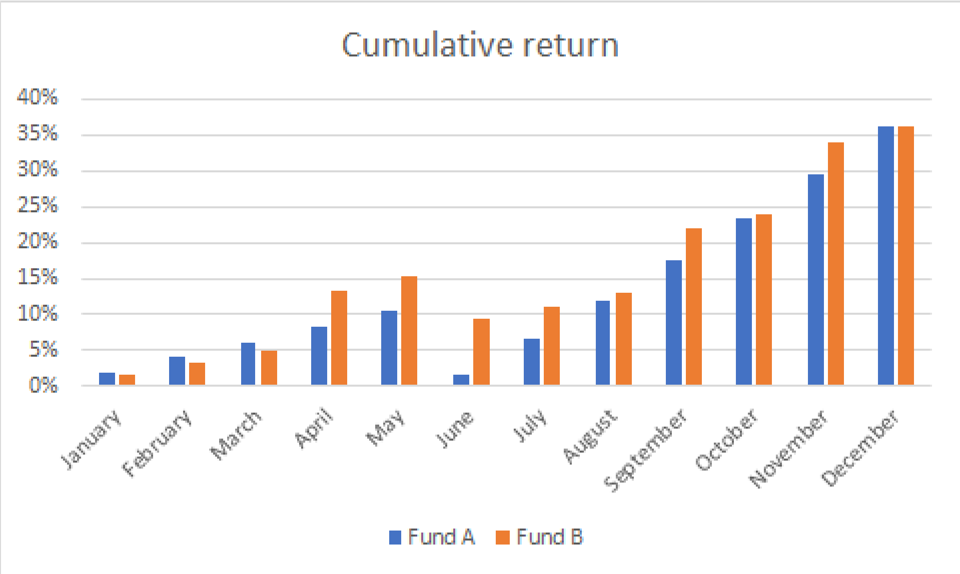
Tversky, A. and Kahneman, D. (1992) 'Advances in prospect theory: Cumulative representation of uncertainty', *Journal of Risk and Uncertainty*. Kluwer Academic Publishers, 5(4), pp. 297–323. doi: 10.1007/BF00122574.

Wermers, R. R. (2003) 'Is Money Really "Smart"? New Evidence on the Relation Between Mutual Fund Flows, Manager Behavior, and Performance Persistence', *SSRN Electronic Journal*. Elsevier BV. doi: 10.2139/ssrn.414420.

Wu, G. and Gonzalez, R. (1996) ‘Curvature of the probability weighting function’, *Management Science*. INFORMS Inst.for Operations Res.and the Management Sciences, 42(12), pp. 1676–1690. doi: 10.1287/mnsc.42.12.1676.

Appendix I – Additional data on the two funds

Figure 5. Graph showing that the two funds being compared have identical cumulative return and volatility.



Appendix II – Summary of variables

Table 5. Summary of variables and their effect

Variable name	**Effect**	**Comment**
LN_AGE	-	Expected effect. Older funds might grow slower compared to new funds.
VOL	+ / -	Volatility could indicate increased risk, as the model adjusts for both adjusted yearly returns and MAX. We would therefore expect a negative effect across regressions.
SKEW	None	Increased skewness implies a longer right tail: meaning investors can expect frequent small losses and fewer large gains. Since we control for adjusted yearly returns and volatility, the skewness can be hard to interpret.
FUND_TNA	-	Expected effect. Larger funds probably grow slower compared to smaller funds.
COMP_TNA	+	Expected effect. Larger fund companies probably have higher visibility and may attract more capital.
LAG_FLOW	+	Expected effect. We expect to see time clusters of FLOW if there are no drastic changes on the market.
MGMT_FEE	+	Unexpected effect. A higher management fee should be unattractive to investors.
TURN_RATIO	None	Insignificant effect and we would perhaps not expect any significant relationship between turnover ratio and fund flow.
EXP_RATIO	-	It is hard to interpret whether a higher expense ratio should have a direct effect on FLOW.

Appendix III - Data filtering

The exact process of data filtering is not described on a detailed level in AG (2020). However, we believe that the process that we have followed is very similar to what they have done in order to achieve their final dataset.

We use the share class codes to download data for the same funds from the CRSP mutual funds database.

Our initial data cleaning removes rows where data for certain information is missing (before merging[29] the datasets for fund information and return data). This removes the following number of rows:

- Turnover ratio - 27,137 rows
- Expense ratio - 27,137 rows
- Age - 1,865 rows
- Management fee - 27,137 rows
- CRSP objective code - 8,497 rows
- Share class code - 4,898 rows
- Management company code - 22,948 rows

After this, we merge the information datasets with the return dataset. We now filter out rows with missing data again, since it is not certain that all rows after the merge contain complete data for all variables. This removes the following amount of rows:

- CRSP objective code - 322,933 rows
- Monthly return (missing value can be denoted by “R”) - 25,894 rows
- Total Net Assets - 4,814 rows
- Share class code - 24,497 rows
- Management code - 167,199 rows
- Objective code - 0 rows
- Management fee - 81,253 rows
- Age - 12 rows
- Expense ratio - 0 rows[30]

[29] CRSP mutual fund data cannot deliver information about the fund and return data in a single file. We therefore must merge datasets containing fund information with the dataset containing return data for the fund.
[30] As we see, the number of missing data points is equal to 27,137 rows for turnover ratio, expense ratio and management fee. It seems like the data is missing for the same rows for these 3 variables and that also explains why 0 rows are removed when filtering out missing expense ratio and turnover ratio; they have already been filtered out with management fee in the merged dataset.

- Turnover ratio - 0 rows

After this, we follow the algorithm that requires each fund to have 15 consecutive months to be included in the dataset. So, every time series with less than 15 consecutive months will be removed. This removes 21,658 rows.

This leads to a dataset that consists of 1,367,867 rows. After all the calculations have been done on a quarterly basis, we are left with 409,663 rows, since we filter out all rows that do not represent the end of quarters, 958,204 rows.

After this, we must remove all rows without lagged variables, since the lagged variables are the variables of primary interest in the regression, removing 14,363 rows. We then filter out the bottom 1-percentile and top 99 percentile of FLOW, just as AG (2020). This is due to the fact that fund mergers and splits might distort the data. This removed 14,495 rows. Finally, we remove all rows with less than $500,000 in AUM, removing 23,129 rows.

After this we remove all rows with expense ratio smaller than 0 (0%) or larger than 1 (100%), removing 24,235 observations. We remove rows with management fee smaller than 0 (0%) or larger than 1 (100%), removing 1,161 observations. We eliminate rows data where turnover ratio is smaller than 0 (0%), removing 1,754 observations.