DUMB AND DUMBER

A STUDY OF CAPITAL FLOWS AND CROSS-SECTIONAL MISPRICING

JOHANNA LINDEGREN
MARCUS HOBER

Bachelor Thesis
Stockholm School of Economics
2021

Dumb and Dumber: A Study of Capital Flows and Cross-sectional Mispricing

Abstract:

This study analyzes the role of *smart money* and *dumb money* in relation to cross-sectional mispricing of stocks, measured using eleven well-documented asset pricing anomalies. Further, we investigate whether *dumber money* is present in the market by examining the relationship between retail investor capital flows and mispricing in the cross section of stocks. We find that mutual fund flows exacerbate aggregate mispricing by buying overvalued stocks, while hedge fund flows attenuate aggregate mispricing. However, the price-correcting effect of hedge fund flows has diminished over time. Finally, our results suggest that retail investor capital flows represent *dumber money*, as the flows exacerbate aggregate mispricing from two ends: both by buying overvalued stocks and by selling undervalued stocks.

Keywords:

Mispricing, Pricing anomalies, Hedge funds, Mutual funds, Retail investors

Authors:

Johanna Lindegren (24409)
Marcus Hober (24476)

Tutor:

Adrien d'Avernas, Assistant Professor, Department of Finance

Examiner:

Adrien d'Avernas, Assistant Professor, Department of Finance

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

I. Introduction

A long-standing interest of economists has been to understand why stock prices may depart from their intrinsic value. The argument against such deviations is that any price divergences from fundamentals should be exploited by sophisticated investors who seek to generate profits. Mispricing is thus expected to vanish as rational investors act on these discrepancies. However, if rational traders cannot fully exploit these profit opportunities, mispricing will prevail, and there is substantial documentation of such pricing anomalies in finance literature. In this paper, we investigate the sources of mispricing. Specifically, we ask which actors exacerbate and attenuate mispricing, thus representing *smart*, *dumb* or perhaps even *dumber money*.

To answer this question, we investigate the relationship between hedge funds, mutual funds, retail investors and cross-sectional mispricing. The first part of our study replicates the paper *Smart Money, Dumb Money and Capital Market Anomalies* by Akbas et al. (2015). They show that mutual fund flows represent *dumb money* that exacerbate cross-sectional mispricing, while hedge fund flows represent *smart money* that attenuate cross-sectional mispricing. This effect is observed in US data between 1994 and 2012, by analyzing the inter-temporal relationship between time series for capital flows and a mispricing metric. Mispricing is measured using the set of anomalies presented by Stambaugh, Yu and Yuan (2012). These anomalies are used to create portfolios that buy undervalued stocks and sell overvalued stocks, thus generating positive returns when cross-sectional mispricing is attenuated, and negative returns when mispricing is corrected.

We extend the study by Akbas et al. (2015) until present day to verify if the proposed relationships between hedge fund flows, mutual fund flows and mispricing prevail. Furthermore, we develop the findings by Akbas et al. (2015) by examining mispricing effects caused by retail investors themselves. We argue that self-directed retail investor capital flows are likely to capture investment decisions taken by naïve investors more precisely. Retail investors play an increasingly important role in capital markets as retail investor capital flows are increasing in size and scope – a development that calls for a better understanding of the implications of increasing stock market participation of these investors. In 2018, there were as many as 2,857 retail broker-dealers in the US, managing 128 million customer accounts (Securities and Exchange Commission (2018)). Behind each of these accounts are individuals who are subject to misconceptions and psychological traps, such as herding and exaggerated media responses (Dalbar (2015)). An intuitive conjecture is that these characteristics are more likely to be prevalent among retail investors than mutual fund managers. We therefore suggest that retail investor capital flows may potentially represent even *dumber money* than mutual fund flows.

Further, retail investors may have a direct impact on stock prices, thus affecting the efficiency of capital allocation. This effect is highly relevant for theoretical purposes, as we can improve the understanding of sources to market inefficiencies. The topic is useful for practical purposes as well, since policy makers may incorporate the findings in the process of regulating equity markets. Democratization of markets is often expected to improve market efficiency, but we investigate if retail investors might even destroy value by making financial markets less efficient. Our paper contributes with insights regarding market efficiency and the sources of cross-sectional mispricing by exploring three hypotheses:

Hypothesis 1: Mutual fund flows exacerbate, and hedge fund flows attenuate cross-sectional mispricing of stocks between 1994 and 2012
Hypothesis 2: Mutual fund flows exacerbate, and hedge fund flows attenuate cross-sectional mispricing of stocks between 1994 and 2020
Hypothesis 3: Retail investor capital flows exacerbate cross-sectional mispricing of stocks

We base the core of our analysis on regressions, where we examine if mutual fund flows, hedge fund flows and retail investor capital flows influence cross-sectional mispricing of stocks, measures as the return of a portfolio with a long position in undervalued stocks and a short position in overvalued stocks. From this analysis, we arrive at three main conclusions: (i) mutual fund flows exacerbate aggregate mispricing, (ii) hedge fund flows attenuate aggregate mispricing, but with a diminishing effect over time and (iii) retail investors increase aggregate mispricing by buying overvalued stocks and selling undervalued stocks. Thus, hedge fund flows represent *smart money*, mutual fund flows *dumb money* and retail investor capital flows even *dumber money* – as retail investors act in contradiction to what is generally seen as a good investment strategy. The fact that retail investors behave not only suboptimally, but in a completely irrational manner, is a valuable indicator that more research is needed to better understand how to design capital markets to improve financial efficiency.

The rest of the paper proceeds as follows. The relevant literature is introduced in Section II, followed by a description of the data and the methodology in Section III. Section IV presents the results and Section V provides robustness checks and corroborative evidence. Finally, Section VI concludes the paper.

II. Literature Review

Our paper extends the literature related to asset pricing and capital flows, where the two closest studies are Akbas et al. (2015) and Stambaugh, Yu and Yuan (2012). Akbas et al. (2015) investigate the effect of capital flows on cross-sectional mispricing, by using the combined set of anomalies presented by Stambaugh, Yu and Yuan (2012). Several articles have analyzed the effect of mutual funds and hedge funds on stock returns, such as Lou (2012), demonstrating that mutual funds may herd and follow momentum strategies. This behavior may have asset-pricing implications, leading stock prices to depart from intrinsic value. As such, mutual fund flows are referred to as *dumb money*, which is a core theme of our paper. Through our replication of Akbas et al. (2015), we strengthen their conclusions, thereby giving further credibility to the empirical results. Moreover, the extension until present day sheds light on how the proposed relationships have evolved over time, which is of particular interest due to rapid changes in the structure of financial markets.

One distinction between Akbas et al. (2015) and our paper relates to the mispricing measures of eleven individual asset pricing anomalies. Akbas et al. (2015) aim to understand the channels through which capital flows influence cross-sectional mispricing, which they do by exploring the relationship between flows and the returns of each of the anomalies. This feature is handled differently in our paper, as we incorporate the mutual conclusion by Akbas et al. (2015) and Stambaugh, Yu and Yuan (2012) that the composite measure of the anomalies seems to be a better predictor of mispricing. We depict the relation between hedge fund flows, mutual fund flows and the return of individual anomalies, but give emphasis to the aggregate mispricing portfolio when steering our focus toward retail investor capital flows.

By using the group of anomalies from Stambaugh, Yu and Yuan (2012), we investigate if the anomalies are equally relevant today. It is possible that changing characteristics – of society in general, and the stock market in particular – could improve or harm market efficiency. Digitalization, faster information flows and a more diverse investor base serve as examples of trends that might have transformed mispricing for the individual anomalies. We deliver corroborative evidence for their relevance, but do not investigate potential strengths or weaknesses of the anomalies, nor suggest alternative measures of mispricing. Moreover, our paper differs from the work of Stambaugh, Yu and Yuan (2012) as we analyze the sources of mispricing rather than observing the mere effect. More specifically, Stambaugh, Yu and Yuan (2012) investigate the presence of cross-sectional mispricing in conjunction with a broad sentiment index constructed by Baker and Wurgler (2006). Instead, the main purpose of our paper is to analyze the direct impact of retail investor actions on mispricing in the cross section of stocks.

We combine the mispricing theme from Stambaugh, Yu and Yuan (2012) with discoveries by Kumar and Lee (2006), who find that retail investors tend to buy the same stocks at the same time, thereby inducing price pressure on individual securities. We further hypothesize that the stock price movements caused by retail investors may lead to mispricing. Thereby, we extend the findings by Kumar and Lee (2006) as we introduce retail investor actions in the setting of mispricing proposed by Stambaugh, Yu and Yuan (2012), since we argue that *dumber money* plays an important role in equity markets.

A key difference between the paper by Kumar and Lee (2006) and ours is that they focus on behavioral patterns among retail investors, rather than aggregate retail investor impact on asset prices. Thereby, we explain the net effect of retail investor activity on cross-sectional mispricing, which enables us to analyze if retail investors on a general level represent *dumber money*. There are no prior studies of the effect of self-directed retail investor capital flows on cross-sectional mispricing, which makes this dimension of our study especially valuable. While our paper illustrates the relationship between capital flows and mispricing, we do not attempt to explain the underlying reasons behind the actions of investors.

Altogether, our paper replicates Akbas et al. (2015) and extends their conclusions with insights from Stambaugh, Yu and Yuan (2012) and Kumar and Lee (2006). By intertwining these three perspectives, we build on previous, influential research and add an increasingly relevant dimension: *dumber money*.

III. Data and Methodology

As we examine the underlying mechanism behind mispricing, we analyze the relationship between hedge funds, mutual funds, retail investors and cross-sectional mispricing of stocks by performing regressions. We use a set of well-documented anomalies as a proxy for cross-sectional mispricing. Further, we collect US data to take advantage of the comprehensive and reliable information accessible for this region, which enables more robust inferences.

A. Variable Descriptions

In this section, we explain the variables used in our regressions. First, we explain how we construct stock portfolios that aim to capture cross-sectional mispricing, which are our dependent variables in the regressions. Thereafter, we outline the construction of our independent variables, *smart money*, *dumb money* and *dumber money*, which are hypothesized to be capital flows from hedge funds, mutual funds and retail investors. Lastly, we introduce our control variables, which are included to account for risk and illiquidity.

A.1. Mispricing Measures

We measure mispricing following the method developed by Stambaugh, Yu and Yuan (2012), which is used by Akbas et al. (2015) as well. By ranking all stocks in our dataset according to eleven individual anomalies, we determine which stocks are most likely to be overvalued and undervalued, respectively. Rankings are assigned each month, and if a stock that is determined to be undervalued generates a positive return in the following month, mispricing is attenuated. Likewise, if the return is negative, mispricing is exacerbated.

Next, all stocks are grouped into deciles based on their rankings, and the extreme deciles are expected to include the most mispriced stocks. By doing this, a hedge portfolio that buys undervalued stocks and sells overvalued stocks is formed. If the strategy correctly captures cross-sectional mispricing, abnormal returns should indicate whether mispricing is exacerbated or attenuated. We create this type of hedge strategy for each anomaly, which enables an assessment of the relationship between capital flows and individual anomalies.

Moreover, Stambaugh, Yu and Yuan (2012) present an important finding; the individual anomalies show weak correlations with each other, while simultaneously showing strong correlations with the returns of an aggregate long-short portfolio based on the composite rankings of the individual anomalies. This result implies that an aggregate mispricing metric is likely to be a more adequate tool to distinguish overvalued and undervalued stocks, as each anomaly seems to capture different aspects of mispricing. Taken together, this should yield a more accurate measure of mispricing than the individual anomaly returns. We investigate this relationship in our empirical analysis. The aggregate portfolio is constructed according to the same procedure as the individual anomalies, but the decile rankings are now used to compute an equal-weighted aggregate score for each stock. Based on these scores, we create new decile rankings used to construct the aggregate long-short portfolio.

The returns of the portfolios based on individual and aggregate anomalies should be interpreted in the same way. During months when mispricing is attenuated, the long-short portfolio produces positive returns. On the contrary, when mispricing is exacerbated, we expect to see negative returns. To construct the anomalies, we use data from the Center for Research in Security Prices (CRSP) and Compustat. The dataset includes stocks listed on NYSE, Amex and NASDAQ. Details on the construction of the anomalies are provided in Appendix A. Next, we describe the intuition behind the anomalies:

Return on assets. Chen, Novy-Marx and Zhang (2011) show that stocks with higher past return on assets earn abnormally higher future returns.

Ohlson O-score. Ohlson (1980) presents a static model that uses accounting data to determine the probability of bankruptcy. Stocks with a high O-score have a higher probability of facing financial distress in the near future, resulting in lower subsequent returns.

Failure probability. The failure probability anomaly is based on a dynamic logit model that explains the negative relationship between failure probability and future stock returns. The model combines a group of equity market and accounting variables to retrieve the probability of distress. The model was developed by Campbell, Hilscher and Szilagyi (2008).

Gross profitability. Novy-Marx (2013) shows that profitability, measured as gross profit-to-assets, is a predictor of the cross section of returns. He argues that gross profitability is the cleanest measure of economic profitability, and that profitability measures become more "polluted" further down the income statement. Stocks with high gross profitability have higher subsequent returns.

Net stock issues. Ritter (1991) and Ritter and Loughran (1995) found that net stock issues and stock returns are negatively correlated, as firm managers tend to issue shares when sentiment-driven investors create upward price pressure on stocks.

Total accrual. Previous research establishes a negative relationship between accruals and cash earnings, whereas firms with high accruals could be engaging in misrepresentative accounting techniques. Sloan (1996) builds on this connection by demonstrating that stocks of firms with low accruals produce higher future stock returns.

Composite equity issues. Daniel and Titman (2006) show that stocks of firms that issue new equity underperform the stocks of firms that do not issue new equity, in line with the findings of Ritter (1991) and Ritter and Loughran (1995).

Investment-to-assets. Titman, Wei and Xie (2004) find that high past investments predict lower subsequent returns, as investors initially underreact to empire-building behavior and overinvestment by firm managers.

Net operating assets. Hirshleifer, Hou, Teoh, and Zhang (2004) find that net operating assets is negatively related to stock returns, as unsophisticated investors value accounting profitability higher than cash conversion.

Asset growth. Cooper, Gulen and Schill (2008) conclude that higher asset growth is related to lower stock returns. This relationship is explained by investors' overreaction to asset expansions and leads to mispricing of growing firms, as investors erroneously expect high-growth firms to perform better.

Momentum. Jegadeesh and Titman (1993) present a momentum effect in the performance of stocks, meaning that stocks tend to continue to move in the same direction as the historical trend. This trend can cause prices to deviate from fundamental value.

A.2. Capital Flows

To analyze the source of mispricing, we introduce proxies for *smart money*, *dumb money* and *dumber money*. We use capital flows to hedge funds, mutual funds and retail investors for this purpose, which are our independent variables. Since we use time series data, a strong time trend is present in the capital flows since the amount of money invested has increased markedly since the beginning of our analysis in 1994. The retail investor customer segment has grown especially fast because of digitalization and increased availability. To account for this effect, our monthly metrics are expressed as percentage changes rather than in absolute amounts.

Hedge fund flows. Capital flows to equity hedge funds is our proxy for *smart money*. The astuteness of the trades performed by hedge funds can be understood through the use of sophisticated compensation schemes and the possibility for short-selling, as presented by Jagannathan, Malakhov and Novikov (2010). These conditions are likely to result in more informed investment decisions, as Boehmer, Jones and Zhang (2008) confirm that short investment decisions are generally well informed, and thereby contribute to efficient pricing. To quantify the measure, aggregate hedge fund flow is computed as

$$HFFLOW_t = \frac{\sum_{i=1}^{N}[HTNA_{i,t} - HTNA_{i,t-1}(1+HRET_{i,t})]}{\sum_{i=1}^{N} HTNA_{i,t-1}}, \tag{1}$$

where $HTNA_{i,t}$ is total net assets for hedge fund i at time t, and $HRET_{i,t}$ is the monthly return for hedge fund i at time t. We define t in months throughout the paper.

We collect hedge fund data from the Lipper Trading Advisor Selection System (TASS) database, spanning from January 1994 until November 2020. The variables obtained are total net assets and returns as well as fund characteristics, in order exclude funds that do not primarily invest in US equities. The Lipper TASS database does not include dead funds prior to 1994 and we have therefore excluded older data to avoid survivorship bias. To get a representative sample for hedge funds primarily trading in US equities, we follow the selection methodology by Cao, Chen, Liang and Lo (2013). Hence, hedge funds in the categories fixed income arbitrage, managed futures and emerging markets are excluded. We also exclude fund-of-funds to avoid double counting. Furthermore, following common practice in the literature, we only include funds with more than USD 10 million in average assets under management. Consequently, our sample includes 275,715 monthly observations.

As noted by Fung and Hsieh (2000), there exists biases in hedge fund databases. These biases are of less concern in our study as we look at aggregate fund flows rather than individual fund performance. On the other hand, reporting requirements are not fully standardized, which may lead to noise in the data. With this in mind, it is of utmost importance to use high-quality data, and we therefore use Lipper TASS, which is considered the leading provider of timely and comprehensive hedge fund information.

Mutual fund flows. As a proxy for *dumb money*, we use capital flows to equity mutual funds. Here, we follow Akbas et al. (2015), who argue that mutual fund flows capture the actions of retail investors in financial markets. Retail investors are the main buyers of mutual funds according to IMF (2014), and previous documentation demonstrates that mutual fund flows affect mispricing. Sirri and Tufano (1998) show that individual investors allocate their money disproportionately to mutual funds with high past performance, while failing to sell funds with lower returns. Further, Frazzini and Lamont (2008) show that retail investors tend to allocate money to mutual funds that buy overvalued stocks. Mutual funds that experience large capital inflows also tend to buy more of the stocks already held, which creates upward price pressure, as concluded by Coval and Stafford (2007). Altogether, research shows that retail investors affect stock prices by allocating capital to mutual funds, whereby mutual fund flows should be a suitable proxy for *dumb money*. The aggregate mutual fund flow measure is computed as

$$MFFLOW_t = \frac{\sum_{i=1}^{N}[MTNA_{i,t} - MTNA_{i,t-1}(1+MRET_{i,t})]}{\sum_{i=1}^{N} MTNA_{i,t-1}}, \tag{2}$$

where $MTNA_{i,t}$ is total net assets for mutual fund i at time t, and $MRET_{i,t}$ is the monthly return, net of fees, for mutual fund i at time t.

Mutual fund flow data is collected from the CRSP Survivor-Bias-Free US Mutual Fund Database. The variables obtained are total net assets and monthly returns for individual mutual funds as well as their fund objectives. We use monthly data from 1994 until 2020 to match the time frame used for the hedge fund flows. Moreover, in line with the method used by Huang, Sialm and Zhang (2011), we only include actively managed mutual funds that primarily invest in US equities and with more than USD 5 million in assets under management in the previous month. This yields a sample of 2,318,496 monthly observations. For more details on the funds included in our sample, see Appendix B.

Retail investor capital flows. Lastly, we use retail investor capital flows as a proxy for *dumber money*. We argue that this is a cleaner measure of retail investor behavior, as individuals make their investment decisions independently instead of through an intermediary. There are few established measures of direct retail investor actions in the stock market, why we create a metric comparable to our proxies for *smart money* and *dumb money*. We base the metric on retail investor capital flows to American stockbrokers, and the aggregate measure is computed as

$$RIFLOW_t = \frac{\sum_{i=1}^{N} RNNA_{i,t}}{\sum_{i=1}^{N} RTNA_{i,t-1}}, \tag{3}$$

where $RNNA_{i,t}$ is net new assets for stockbroker i at time t, and $RTNA_{i,t}$ is total net assets for stockbroker i at time t. Retail investor capital flow data is collected from monthly reports on trading activity from Charles Schwab and E*TRADE. The reports from Charles Schwab are retrieved from the US Securities and Exchange Commission's Archive of historical EDGAR documents. The reports from E*TRADE are retrieved from their press release library. The sample covers the full time period where data is accessible, which is between April 2008 and December 2020 for Charles Schwab and between April 2008 and August 2020 for E*TRADE.[1] Data from other US brokerage firms is not publicly available.

A.3. Fama-French Three Factors

Our first three control variables are obtained from the three-factor model developed by Fama and French (1993). The model is used to explain stock returns and includes the excess return of the market, the size factor and the value factor. Although there is an ongoing debate whether these three factors represent risk or mispricing – for example, as discussed by Bloomfield and Michaely (2004) – the model is widely used to control for risk. The data is accessed through Wharton Research Data Services (WRDS).

A.4. Market Illiquidity

Our last two control variables are used to account for the well-documented effect of liquidity on stock prices. For example, Silber (1975) demonstrates that a thin market experiences large changes in prices from small changes in supply or demand. Thus, we expect the return of the long-short hedge strategy to be higher when the market is liquid, since it is easier for investors to engage in price-correcting trades. Therefore, it is necessary to control for this phenomenon in our regressions. There are plenty of measures of illiquidity, such as the bid-ask spread used by Amihud and Mendelson (1986) and the dollar volume of trading used by Brennan, Chordia and Subrahmanyam (1998). Amihud (2002) states that it is unlikely that one single measure can be created to capture all facets of illiquidity and we therefore use two different control variables to account for illiquidity, in line with Akbas et al. (2015).

The first illiquidity measure is aggregate illiquidity, which is computed according to the definition by Amihud (2002). Amihud (2002) defines illiquidity as the average ratio of the daily absolute return to the daily trading volume of stocks. This is one of the most commonly used proxies for illiquidity, and a high measure indicates a larger price impact of trade, and should be associated with less correction of aggregate mispricing. To compute the metric, we use daily data from CRSP. We exclude shares on other exchanges than NYSE, with a share price lower than USD 5 in the previous month and with less than 200 days of observations in the previous year. Furthermore, the highest and lowest percentiles each month are excluded.

Our second measure of illiquidity is aggregate turnover. According to Barinov (2014), finance literature has used turnover – specifically trading volume over shares outstanding – as a proxy for illiquidity for a long time. Aggregate turnover captures that investors can trade in stocks at low cost when turnover is high. In our paper, we use the illiquidity turnover metric proposed by Datar, Naik and Radcliffe (1998), where trading volume is divided by shares outstanding. When aggregate turnover is high, it is easier to trade and thereby easier to correct mispricing. Thus, correction of mispricing should be more prominent in months with high turnover. To compute aggregate turnover, we use monthly data

[1] Since we compute the percentage change in retail investor capital flows, our regressions include 152 monthly observations, from May 2008 until December 2020.

from CRSP. We only include stocks listed on NYSE and with a share price higher than USD 5 in the previous month.

Stocks listed on NYSE are analyzed in isolation for two reasons. Firstly, trading volume is measured differently on NYSE and NASDAQ. NYSE functions as an auction market while NASDAQ is a dealer market, using many competing market makers. Secondly, there are differences in liquidity premiums between the exchanges and Reinganum (1990) finds that NASDAQ has a liquidity advantage for smaller firms. Since NYSE is the largest stock exchange with a more diverse list of traded companies, it is the appropriate choice for our research purpose.

B. Statistical Model

We conduct our analysis by performing several regressions. Firstly, we use regressions to assess the performance of the long-short portfolios as measures of cross-sectional mispricing. We use the Fama-French three-factor model for this purpose, with the dependent variable being the portfolio returns. The use of the Fama-French model sets a higher performance hurdle for the portfolio returns compared to the capital asset pricing model by Sharpe (1964) and Lintner (1965). This way, we can verify a more reliable set of anomalies. The first regression equation looks as follows.

$$R_{p,t} = \alpha_p + \beta_1 \times RMRF_t + \beta_2 \times SMB_t + \beta_3 \times HML_t + \epsilon_{p,t} \tag{4}$$

A regression that results in a significant alpha indicates that the constructed portfolio strategy captures mispricing. Once a mispricing metric is validated, we can use it to assess the relationship between capital flows and mispricing, which is the second purpose of our use of regressions.

We perform one type of regression to examine our first and second hypotheses, and another regression to investigate our third hypothesis. The first two hypotheses aim to analyze the relation between mutual fund flows and hedge fund flows and the portfolio returns of individual and aggregate anomalies. The predictive power is controlled for by including the market, size and value factors, as well as the illiquidity measures. Altogether, the regression equation is stated below.

$$\begin{aligned} R_{p,t} = \alpha_p + \beta_1 \times MFFLOW_t + \beta_2 \times HFFLOW_t + \beta_3 \times AGGILLIQ_t + \beta_4 \times AGGTURN_t \\ + \beta_5 \times RMRF_t + \beta_6 \times SMB_t + \beta_7 \times HML_t + \epsilon_{p,t} \end{aligned} \tag{5}$$

To test our third hypothesis – that retail investor capital flows exacerbate mispricing – we switch the independent variables to retail investor capital flows.

$$\begin{aligned} R_{p,t} = \alpha_p + \beta_1 \times RIFLOW_t + \beta_2 \times AGGILLIQ_t + \beta_3 \times AGGTURN_t + \beta_4 \times RMRF_t \\ + \beta_5 \times SMB_t + \beta_6 \times HML_t + \epsilon_{p,t} \end{aligned} \tag{6}$$

As we use time series data in our regressions, there is a natural risk for heteroscedasticity and autocorrelation in the error terms, which might lead to inaccurate predictions.

For this reason, we test our data for heteroscedasticity by performing Breusch-Pagan tests (Breusch and Pagan (1979)), which affirms that our residuals are heteroscedastic. We continue by examining whether our data shows signs of autocorrelation with Durbin-Watson tests (Durbin and Watson (1950), Durbin and Watson (1951)), and find that autocorrelation is present in our samples as well. The presence of heteroscedasticity and autocorrelation calls for an adjustment to our model, why we use Newey-West standard errors to compute the t-values throughout our empirical analysis.

IV. Results

This section presents the results of the empirical analysis conducted to test our hypotheses. The analysis consists of three parts: (i) our replication of Akbas et al. (2015), testing our first hypothesis, (ii) our extension until present day, testing our second hypothesis and (iii) our investigation of retail investor capital flows, testing our third hypothesis.

A. *Mutual Fund Flows Exacerbate, and Hedge Fund Flows Attenuate Cross-sectional Mispricing of Stocks between 1994 and 2012*

We begin by presenting characteristics of the key variables used to test the first hypothesis, which are outlined in Table I. The univariate statistics of the long-short portfolio, seen in Panel A, give an initial indication that our long-short portfolio is an accurate measure of mispricing, since the mean excess return is positive (1.52%). Looking at the long leg separately, the portfolio performs as expected with an average monthly excess return of 1.14%. Further, the short leg is expected to produce negative returns, since the stocks held in this portfolio are determined to be overvalued. This prediction is confirmed as the short portfolio has an average return that is even lower than the T-bill rate, with a negative average monthly excess return (-0.38%). These results, together with the standard deviation and the sample size, reveal that the return of the long-short strategy is significantly different from zero, indicating that the portfolio successfully captures aggregate mispricing by distinguishing over- and undervalued stocks.[2]

Next, we look at the correlations between our variables, which are presented in Panel B of Table I. To begin with, we notice a positive correlation between mutual fund flows and hedge fund flows ($\rho = 0.187, p = 0.00$). Even though this result is significant, the magnitude of the correlation is sufficiently low to allow for inter-temporal variation in the comovements of the two variables, where the sign of the correlation may vary from month to month. This variation is important to enable meaningful interpretations of the results that follow, as we can distinguish the effect that each variable has on the left-hand side of the equation.

Moreover, mutual fund flows are significantly, negatively correlated with the long-short portfolio ($\rho = 0.239,\ p = 0.00$), while hedge fund flows are positively correlated with the portfolio ($\rho = 0.088,\ p = 0.19$). This gives a hint that our first hypothesis may be accurate in the respect that mutual fund flows exacerbate mispricing and hedge fund flows attenuate mis-

[2] The distributions of these variables are similar to those presented by Akbas et al. (2015), see Appendix E.

Table I
Summary Statistics, 1994-2012

This table reports univariate summary statistics from January 1994 to December 2012 for all variables used to test our first hypothesis. The flow variables are the monthly mean of aggregate equity mutual fund flows (MFFLOW) and equity hedge fund flows (HFFLOW). LONG and SHORT represent the monthly excess return series of the long and short leg in the mispricing metric (L-S), constructed from the eleven anomalies described in Section III. Control variables are monthly excess market returns (RMRF), aggregate illiquidity (AGGILLIQ), aggregate turnover (AGGTURN), the value factor (HML) and the size factor (SMB). Panel B shows the Pearson pairwise correlation estimates. P-values are listed below each estimate.

Panel A: Descriptive Statistics

Variable	N	Mean	Median	Standard deviation	Minimum	10th percentile	25th percentile	75th percentile	90th percentile	Maximum
MFFLOW	228	0.0034	0.0032	0.005	-0.013	-0.002	0.000	0.006	0.010	0.022
HFFLOW	228	0.0050	0.0071	0.017	-0.108	-0.011	-0.003	0.015	0.022	0.039
LONG	228	0.0114	0.0170	0.045	-0.166	-0.051	-0.014	0.041	0.060	0.134
SHORT	228	-0.0038	0.0041	0.072	-0.264	-0.095	-0.046	0.042	0.078	0.234
L-S	228	0.0152	0.0144	0.040	-0.159	-0.025	-0.003	0.033	0.057	0.143
RMRF	228	0.0052	0.0119	0.046	-0.172	-0.056	-0.022	0.035	0.061	0.114
AGGILLIQ	228	0.0357	0.0341	0.023	0.007	0.009	0.013	0.055	0.067	0.090
AGGTURN	228	0.1303	0.1129	0.062	0.052	0.066	0.075	0.182	0.219	0.359
HML	228	0.0027	0.0014	0.032	-0.111	-0.032	-0.015	0.018	0.036	0.126
SMB	228	0.0015	0.0005	0.035	-0.168	-0.036	-0.019	0.021	0.038	0.212

Panel B: Pairwise Correlations

Variable	MFFLOW	HFFLOW	LONG	SHORT	L-S	RMRF	AGGILLIQ	AGGTURN	HML
HFFLOW	0.187								
	0.00								
LONG	0.343	0.074							
	0.00	0.27							
SHORT	0.345	-0.003	0.868						
	0.00	0.96	0.00						
L-S	-0.239	0.088	-0.448	-0.833					
	0.00	0.19	0.00	0.00					
RMRF	0.354	0.022	0.887	0.864	-0.567				
	0.00	0.74	0.00	0.00	0.00				
AGGILLIQ	0.367	-0.053	-0.030	-0.143	0.224	-0.079			
	0.00	0.43	0.65	0.03	0.00	0.23			
AGGTURN	-0.501	-0.284	-0.092	0.025	-0.147	-0.038	-0.680		
	0.00	0.00	0.17	0.71	0.03	0.57	0.00		
HML	-0.040	0.161	-0.096	-0.195	0.244	-0.170	0.059	-0.108	
	0.55	0.02	0.15	0.00	0.00	0.01	0.38	0.10	
SMB	0.125	0.015	0.508	0.536	-0.399	0.222	-0.092	0.037	-0.348
	0.06	0.83	0.00	0.00	0.00	0.00	0.17	0.57	0.00

pricing. Further, our mispricing measure displays a significant negative correlation with the excess return of the market ($\rho = -0.567,\ p = 0.00$) and the size factor ($\rho = -0.399,\ p = 0.19$), suggesting that price-corrections tend to occur in bear markets and when large stocks outperform small stocks. Perhaps, bear markets force investors to identify mispriced stocks, as they cannot rely on the overall rise of the market to generate returns. We also notice a positive relation between the returns of the long-short portfolio and the value factor ($\rho = 0.244,\ p = 0.00$). When looking at market illiquidity, the hedge strategy is positively correlated to aggregate illiquidity ($\rho = 0.244,\ p = 0.00$) and negatively correlated with aggregate turnover ($\rho = -0.147,\ p = 0.03$). These relations suggest that mispricing is attenuated when markets are less liquid. The results are somewhat surprising, as we would expect price-correction to be more prominent when it is easier to trade.[3] Nevertheless, we include all five control variables in our regressions, as the correlations are significant.[4]

In Table II, we present the performance of the hedge portfolios that represent our mispricing metrics. For completeness, the mean excess returns of the aggregate long-short strategy and the long and short components are repeated on the left-hand side of Panel A, together with the t-statistics. As mentioned, the returns of the long-short portfolio are significantly different from zero ($t = 4.97$). On the right-hand side, we include the Fama-French factors in the regressions, using the same dependent variables. The alphas are highly significant, indicating that the portfolios successfully capture aggregate mispricing. The economic magnitude of the alphas further shows that most of the profits of the long-short portfolio are attributable to the short leg (-1.18% versus 0.59%). The difference between the alphas implies that overvalued stocks exhibit a higher degree of mispricing than undervalued stocks.

Further, we examine the returns of the long-short portfolios based on the individual anomalies, which are illustrated in Panel B of Table II. We run the regressions together with the same Fama-French factors but exclude these coefficients in the table for brevity. All intercepts are positive and most of them are highly significant, which suggests that most of the anomalies are good measures of cross-sectional mispricing.

Our results raise a question regarding why the portfolios generate abnormal returns. In theory, sophisticated investors are expected to instantly seize any arbitrage opportunities, reducing the alphas of our portfolios to zero. On the contrary, our findings reveal that cross-sectional mispricing seems to persist over relatively long periods of time. As confirmed by previous literature exploring limits to arbitrage, actions to correct mispricing may require both risk-taking and capital. For example, Shleifer and Vishny (1997) demonstrate that professional arbitrageurs avoid positions that are extremely volatile, and Abreu and Brunnermeier (2002) show that arbitrage trades are delayed because investors await the timing of other arbitrageurs' actions to minimize holding costs. These and other potential limits to arbitrage are likely contributors to our results.

As we now have confirmed that our long-short portfolios successfully capture mispricing, we continue by examining the relation between mispricing and capital flows. To begin with, we assess this relationship using mispricing measures based on the individual anomalies, which are depicted in Table III. Considering the first hypothesis, we anticipate mutual fund

[3] Our results are consistent with those of Akbas et al. (2015).

[4] The signs and the magnitudes of the mentioned correlations show no noteworthy differences from those of Akbas et al. (2015), see Appendix E.

Table II

Mispricing Metric: Returns to a Long-Short Strategy that Uses Cross-Sectional Return Predictors, 1994-2012

The left-hand side of Panel A highlights the mean excess returns of a long-short portfolio (L-S), which is our cross-sectional strategy constructed to capture mispricing, and the long and short components (LONG and SHORT). The right-hand side of Panel A shows Fama-French three-factor alphas for the cross-sectional investment strategy (L-S) and its two legs (LONG and SHORT) in the regression:

$$R_{p,t} = \alpha_p + \beta_1 \times RMRF_t + \beta_2 \times SMB_t + \beta_3 \times HML_t + \epsilon_{p,t}$$

Panel B shows the intercepts from regressions of eleven individual anomalies that are the basis for the long-short portfolio. The anomalies are return on assets (ROA), Ohlson O-score (OO), failure probability (FP), gross profitability (GP), net stock issues (NSI), total accrual (TA), composite equity issues (CEI), investment-to-assets (ITA), net operating assets (NOA), asset growth (AG) and momentum (M). The same Fama-French three-factor model is used. For brevity, the coefficients for the three factors are not tabulated. The t-statistics are listed below the coefficient estimates and are based on Newey-West standard errors.

Panel A: Mispricing Metric Returns

	Mean excess returns			Fama French three-factor alphas		
Variable	L-S	LONG	SHORT	L-S	LONG	SHORT
Alpha	0.0152	0.0114	-0.0038	0.0176	0.0058	-0.0118
	4.97	3.39	-0.69	8.44	5.86	-6.68
RMRF				-0.4348	0.8110	1.2459
				-5.99	24.03	20.34
HML				0.0833	0.2492	0.1659
				0.70	3.60	1.52
SMB				-0.3044	0.4920	0.7964
				-5.27	8.30	7.97
N	228	228	228	228	228	228
Adj. R^2				0.396	0.915	0.874

Panel B: Individual Anomaly Returns

	ROA	OO	FP	GP	NSI	TA	CEI	ITA	NOA	AG	M
FF3 alphas	0.0103	0.0067	0.0151	0.0088	0.0078	0.0023	0.0070	0.0037	0.0078	0.0030	0.0131
	4.46	3.71	4.91	4.38	4.42	1.05	3.76	1.70	3.74	1.62	2.89
N	228	228	228	228	228	228	228	228	228	228	228
Adj. R^2	0.468	0.436	0.433	0.382	0.349	0.049	0.452	0.025	0.032	0.395	0.118

flows to be negatively related to the returns of the long-short strategies, since we expect net capital inflows to result in lower contemporaneous returns of the portfolios. As predicted, all coefficients for mutual fund flows are negative and thus indicate that mutual fund flows exacerbate cross-sectional mispricing across all eleven anomalies. Furthermore, we expect hedge fund flows and the mispricing metrics to display the opposite relationship. This conjecture is supported as the majority of the coefficients for hedge fund flows are positive. Most of the t-statistics for mutual fund flows and hedge fund flows are insignificant, but we are not surprised by these results. While we would expect the positive relation between mutual fund flows and mispricing to be unintentional, our hypothesis suggests that hedge fund managers, representing *smart money*, should target mispriced stocks intentionally to generate profits. This type of trading requires a precise mispricing signal. We acknowledge that there is a significant amount of noise in the individual anomaly returns, and hedge fund managers should therefore incorporate a wide range of known anomalies in their trading strategies, to achieve a higher risk-adjusted return.

To examine this conjecture, we compute Sharpe ratios for the long-short portfolios and compare the ratio of the aggregate mispricing metric with the ratios of the individual anomalies. We find that the Sharpe ratio is consistently higher for the aggregate measure, and the results are highly significant for all portfolios except for one (see Appendix D). To incorporate these findings in our analysis, we rely on the aggregate portfolio to measure mispricing, as it is reasonable to believe that hedge fund managers trade on composite signals, rather than on individual anomalies.[5] This approach to measure cross-sectional mispricing is further supported by Stambaugh, Yu and Yuan (2012), who argue that an aggregate portfolio constructed using several cross-sectional return predictors is a more precise measure, as several dimensions of mispricing are captured in the metric. Also, the high level of noise in the mispricing measures based on individual anomalies is diversified away. To incorporate these findings, we continue by examining the relationship between aggregate mispricing and capital flows.

In Table IV, we repeat the same type of regressions as in Table III, but we let the aggregate measure of mispricing be the dependent variable. Here, our main focus is aimed at the long-short portfolio, which is depicted in the first column of the table. We also present the long and short legs of the strategy separately to allow for a more granular understanding of potential drivers of mispricing. In the light of our first hypothesis, we would expect mutual fund flows to be negatively related to the long-short portfolio returns. This is precisely what we observe in the first column of the table. The significantly negative coefficient ($-1.392, t = -3.69$) suggests that mutual fund flows exacerbate aggregate mispricing in the cross section. By aggregating the anomaly rankings in this way, we obtain a more precise mispricing measure that enables robust inferences, compared to the regressions presented in Table III. In addition, by examining the long and short legs separately, we can understand whether mispricing is mainly affected through transactions in overvalued or undervalued stocks. We find that mutual fund flows are significantly positively related to the short component of the mispricing portfolio ($1.056, t = 3.32$), while the relation to the long leg is less significant and notably smaller in magnitude. This relationship implies that mutual fund

[5] Akbas et al. (2015) obtain similar results, where the majority of the coefficients for mutual fund flows are negative, and the coefficents for hedge fund flows are mixed, see Appendix E.

Table III
Aggregate Mutual Fund Flows, Hedge Fund Flows and Individual Anomaly Returns, 1994-2012

This table shows the coefficients of time series regressions between January 1994 and December 2012, with the return of the long-short strategy (L-S) for eleven individual anomalies as the dependent variable. The anomalies are return on assets (ROA), Ohlson O-score (OO), failure probability (FP), gross profitability (GP), net stock issues (NSI), total accrual (TA), composite equity issues (CEI), investment-to-assets (ITA), net operating assets (NOA), asset growth (AG) and momentum (M). The independent variables are aggregate mutual fund flows (MFFLOW) and aggregate hedge fund flows (HFFLOW). Control variables are aggregate illiquidity (AGGILLIQ), aggregate turnover (AGGTURN), the excess return of the market (RMRF), the size factor (SMB) and the value factor (HML).

$$R_{p,t} = \alpha_t + \beta_1 \times MFFLOW_t + \beta_2 \times HFFLOW_t + \beta_3 \times AGGILLIQ_t + \beta_4 \times AGGTURN_t + \beta_5 \times RMRF_t + \beta_6 \times SMB_t + \beta_7 \times HML_t + \epsilon_{p,t}$$

The t-statistics are listed below the coefficient estimates and are based on Newey-West standard errors.

	ROA	OO	FP	GP	NSI	TA	CEI	ITA	NOA	AG	M
MFFLOW	-0.229	-0.678	-0.479	-0.160	-0.534	-0.313	-0.271	-0.431	-0.680	-0.296	-1.110
	-0.31	-1.56	-0.42	-0.25	-1.47	-0.73	-0.76	-0.77	-1.72	-0.70	-1.00
HFFLOW	0.227	0.160	0.219	0.103	0.004	0.095	-0.008	-0.012	0.037	-0.099	0.540
	1.65	1.00	0.83	0.61	0.04	0.83	-0.08	-0.08	0.29	-0.77	1.33
AGGILLIQ	0.250	0.311	0.041	0.186	0.188	0.111	0.194	0.195	0.331	0.183	-0.278
	1.40	2.92	0.15	1.09	1.58	0.94	1.60	1.77	3.75	1.72	-1.10
AGGTURN	0.076	-0.002	-0.013	0.048	0.010	-0.023	0.024	-0.014	-0.018	0.010	-0.246
	1.24	-0.04	-0.17	0.80	0.21	-0.54	0.52	-0.33	-0.44	0.24	-2.08
RMRF	-0.466	-0.087	-1.066	-0.429	-0.294	0.093	-0.283	-0.066	0.107	-0.189	-0.518
	-6.21	-1.39	-9.03	-7.61	-5.22	1.33	-5.27	-0.93	2.14	-2.41	-2.94
HML	-0.021	-0.476	-0.326	-0.607	0.124	-0.040	0.353	0.138	0.172	0.711	-0.639
	-0.18	-7.55	-1.06	-4.23	1.19	-0.44	3.32	1.52	2.02	5.15	-1.84
SMB	-0.564	-0.674	-0.242	-0.020	-0.203	-0.237	-0.251	0.099	-0.010	0.243	0.213
	-6.53	-8.15	-1.17	-0.32	-3.05	-3.40	-3.76	1.15	-0.08	1.89	0.82
Intercept	-0.009	-0.003	0.016	-0.004	0.001	0.002	-0.002	0.000	0.000	-0.003	0.056
	-0.71	-0.33	1.01	-0.33	0.14	0.21	-0.22	0.00	0.03	-0.37	2.76
N	228	228	228	228	228	228	228	228	228	228	228
Adj. R^2	0.468	0.459	0.425	0.376	0.351	0.044	0.451	0.030	0.075	0.397	0.145

Table IV
Aggregate Mutual Fund Flows, Hedge Fund Flows and Cross-sectional Mispricing, 1994-2012

This table shows the coefficients of time series regressions between January 1994 and December 2012, with the return series of the long-short strategy (L-S), the long leg (LONG) and the short leg (SHORT) as the dependent variables. The portfolios are constructed from the aggregate measure of eleven anomalies. The independent variables are aggregate mutual fund flows (MFFLOW) and hedge fund flows (HFFLOW). Control variables are aggregate illiquidity (AGGILLIQ), aggregate turnover (AGGTURN), the excess return of the market (RMRF), the size factor (SMB) and the value factor (HML).

$$R_{p,t} = \alpha_t + \beta_1 \times MFFLOW_t + \beta_2 \times HFFLOW_t + \beta_3 \times AGGILLIQ_t + \beta_4 \times AGGTURN_t + \beta_5 \times RMRF_t + \beta_6 \times SMB_t + \beta_7 \times HML_t + \epsilon_{p,t}$$

The t-statistics are listed below the coefficient estimates and are based on Newey-West standard errors.

	Mispricing Metric		
Variable	L-S	LONG	SHORT
MFFLOW	-1.392	-0.336	1.056
	-3.69	-1.74	3.32
HFFLOW	0.265	0.060	-0.204
	2.15	0.77	-1.91
AGGILLIQ	0.310	0.112	-0.199
	2.67	2.10	-2.05
AGGTURN	-0.057	-0.022	0.035
	-1.07	-1.00	0.87
RMRF	-0.378	0.825	1.203
	-4.81	25.71	20.21
HML	0.047	0.239	0.191
	0.35	4.28	2.01
SMB	-0.285	0.499	0.784
	-4.84	9.30	7.50
Intercept	0.017	0.005	-0.012
	1.66	1.28	-1.55
N	228	228	228
Adj. R^2	0.443	0.919	0.880

flows primarily exacerbate mispricing of overvalued stocks, as the return of the short portfolio is positive when capital flows of mutual funds enter the market, and vice versa. The result is in line with findings by Miller (1977), who demonstrates that short-sale impediments limit investors' ability to engage in price-correcting trades of overpriced stocks.

We now turn to smart money, for which the coefficients are provided in the second row of Table IV. Hedge fund flows show a significantly positive relationship with the long-short portfolio ($0.265, t = 2.15$), implying that hedge fund flows attenuate mispricing on an aggregate level. Further, hedge fund flows are negatively related to the short leg, which implies that the price-correction primarily occurs through short positions in overvalued stocks. When comparing these results with the ambiguous signals of the individual anomalies, the difference suggests that hedge fund flows are truly *smart money*. Hedge fund managers do not trade on individual anomalies, but rather on the aggregate signal from all anomalies combined, as this strategy enables the highest risk-adjusted return.[6]

While not crucial for our main analysis, it is interesting to note that the economic magnitude of the coefficient for mutual fund flows is large relative to that of hedge fund flows. Thereby, it appears as if mutual fund flows have a more prominent effect on mispricing. This difference is intuitive, as the assets under management of mutual funds have a greater monetary value compared to those of hedge funds; the size of the mutual fund industry is approximately 15 times larger than the hedge fund industry (Akbas et al. (2015)).

To conclude this section, the results corroborate our first hypothesis that mutual fund flows exacerbate mispricing and hedge fund flows attenuate mispricing.

B. Mutual Fund Flows Exacerbate, and Hedge Fund Flows Attenuate Cross-sectional Mispricing of Stocks between 1994 and 2020

We now continue by extending the empirical analysis presented above until 2020 to investigate whether the relationships between fund flows and mispricing prevail.
We begin by looking at the correlations between the regression variables, presented in Table V. While most of the correlation estimates essentially remain unchanged compared to the period 1994 to 2012, a few key characteristics are worth pointing out. First of all, the long-short portfolio is still significantly correlated with the market, size and value factors, why we continue to control for these relations in the following regressions. Moreover, the relationships with the illiquidity measures persist as well.

We continue by validating the mispricing metrics for the longer time period. As outlined on the left-hand side of Panel A of Table VI, the mean return of the long-short strategy is 1.04%, which is smaller than the previous return of 1.52%. However, the return is still significantly positive (t = 4.02). The difference is primarily derived from a higher excess return of the short leg (0.01% instead of -0.38%) as the returns of the long leg are close to identical (1.16% versus 1.14%). More importantly, when controlling for the market, size and value factors, the long-short portfolio generates a significantly positive alpha, which is presented on the right-hand side of Panel A of Table VI. Both legs of the portfolio are significantly different from zero, with an alpha of 0.52% for the long portfolio and -0.82% for the short portfolio. These monthly returns imply that the portfolios continue to successfully capture both undervalued and overvalued stocks over the extended timeframe.

While the portfolio returns of the individual anomalies are of less importance for our main results, it is valuable to understand the constituents of the aggregate long-short portfolio. As shown in Panel B of Table VI, we investigate the alphas of the long-short portfolios based on the individual anomalies, using the Fama-French three-factor model. All alphas continue to be significantly positive, except for two anomalies. The principal point is that the anomalies still capture mispricing on an aggregate level.[7]

The main results of our second hypothesis are provided in Table VII, where we investigate the relationship between the aggregate mispricing measure and fund flows. Recall that,

[6] The results of the regressions allow us to draw the same conclusions as Akbas et al. (2015) regarding mutual and hedge fund flows' impact on aggregate mispricing, see Appendix E.

[7] In untabulated tests, we investigate the relationship between capital flows and the individual anomalies and obtain similar results as those presented in Panel B of Table III.

Table V
Pairwise Correlations, 1994-2020

This table shows Pearson pairwise correlation estimates from January 1994 to November 2020 for all variables used to test our second hypothesis. The flow variables are the monthly mean of aggregate equity mutual fund flows (MFFLOW) and equity hedge fund flows (HFFLOW). LONG and SHORT represent the monthly excess return series of the long and short legs in the mispricing metric (L-S), constructed from the eleven anomalies described in Section III. Control variables are monthly excess market returns (RMRF), aggregate illiquidity (AGGILLIQ), aggregate turnover (AGGTURN), the value factor (HML) and the size factor (SMB). P-values are listed below each estimate.

Variable	MFFLOW	HFFLOW	LONG	SHORT	L-S	RMRF	AGGILLIQ	AGGTURN	HML
HFFLOW	0.227								
	0.00								
LONG	0.311	0.059							
	0.00	0.29							
SHORT	0.266	0.000	0.856						
	0.00	1.00	0.00						
L-S	-0.118	0.068	-0.385	-0.806					
	0.03	0.22	0.00	0.00					
RMRF	0.292	0.009	0.905	0.864	-0.508				
	0.00	0.88	0.00	0.00	0.00				
AGGILLIQ	0.449	0.042	-0.030	-0.171	0.272	-0.102			
	0.00	0.45	0.59	0.00	0.00	0.07			
AGGTURN	-0.540	-0.308	-0.041	0.095	-0.217	0.032	-0.705		
	0.00	0.00	0.46	0.09	0.00	0.57	0.00		
HML	0.041	0.164	-0.019	-0.131	0.212	-0.083	0.098	-0.150	
	0.47	0.00	0.73	0.02	0.00	0.14	0.08	0.01	
SMB	0.109	0.018	0.490	0.527	-0.381	0.246	-0.058	0.045	-0.231
	0.05	0.75	0.00	0.00	0.00	0.00	0.30	0.42	0.00

Table VI
Mispricing Metric: Returns to a Long-Short Strategy that Uses Cross-Sectional Return Predictors, 1994-2020

This table is equivalent to Table II, but covers the entire time frame between January 1994 and November 2020. The left-hand side of Panel A highlights the mean excess returns of a long-short portfolio (L-S), which is our cross-sectional strategy constructed to capture mispricing, and the long and short components (LONG and SHORT). The right-hand side of Panel A shows Fama-French three-factor alphas for the cross-sectional investment strategy (L-S) and its two legs (LONG and SHORT) in the regression:

$$R_{p,t} = \alpha_p + \beta_1 \times RMRF_t + \beta_2 \times SMB_t + \beta_3 \times HML_t + \epsilon_{p,t}$$

Panel B shows the intercepts from regressions of eleven individual anomalies that are the basis for the long-short portfolio. The anomalies are return on assets (ROA), Ohlson O-score (OO), failure probability (FP), gross profitability (GP), net stock issues (NSI), total accrual (TA), composite equity issues (CEI), investment-to-assets (ITA), net operating assets (NOA), asset growth (AG) and momentum (M). The same Fama-French three-factor model is used. For brevity, the coefficients for the three factors are not tabulated. The t-statistics are listed below the coefficient estimates and are based on Newey-West standard errors.

Panel A: Mispricing Metric Returns

	Mean excess returns			Fama French three-factor alphas		
Variable	L-S	LONG	SHORT	L-S	LONG	SHORT
Alpha	0.0104	0.0116	0.0012	0.0134	0.0052	-0.0082
	4.02	4.68	0.29	6.60	5.08	-4.88
RMRF				-0.3767	0.8268	1.2035
				-5.50	29.47	23.31
HML				0.1452	0.1728	0.0275
				1.27	2.78	0.38
SMB				-0.2931	0.4262	0.7193
				-4.63	8.99	8.88
N	323	323	323	323	323	323
Adj. R^2				0.335	0.908	0.850

Panel B: Individual Anomaly Returns

	ROA	OO	FP	GP	NSI	TA	CEI	ITA	NOA	AG	M
FF3 Alphas	0.0096	0.0037	0.0150	0.0069	0.0073	-0.0011	0.0058	0.0024	0.0082	0.0037	0.0131
	4.88	2.34	4.73	3.49	4.93	-0.59	3.63	1.35	4.81	2.17	3.89
N	323	323	323	323	323	323	323	323	323	323	323
Adj. R^2	0.473	0.371	0.450	0.300	0.324	0.047	0.394	0.059	0.024	0.324	0.169

Table VII
Aggregate Mutual Fund Flows, Hedge Fund Flows and Cross-Sectional Mispricing, 1994-2020

This table is equivalent to Table IV, but covers the entire time frame from January 1994 to November 2020. The table shows the coefficients of time series regressions with the return series of the long-short strategy (L-S), the long leg (LONG) and the short leg (SHORT) as the dependent variables. The portfolios are constructed from the aggregate measure of eleven anomalies. The independent variables are aggregate mutual fund flows (MFFLOW) and aggregate hedge fund flows (HFFLOW). Control variables are aggregate illiquidity (AGGILLIQ), aggregate turnover (AGGTURN), the excess return of the market (RMRF), the size factor (SMB) and the value factor (HML).

$$R_{p,t} = \alpha_t + \beta_1 \times MFFLOW_t + \beta_2 \times HFFLOW_t + \beta_3 \times AGGILLIQ_t + \beta_4 \times AGGTURN_t + \beta_5 \times RMRF_t + \beta_6 \times SMB_t + \beta_7 \times HML_t + \epsilon_{p,t}$$

The t-statistics are listed below the coefficient estimates and are based on Newey-West standard errors.

	Mispricing metric		
Variable	L-S	LONG	SHORT
MFFLOW	-1.088	-0.187	0.901
	-3.32	-0.95	3.05
HFFLOW	0.135	0.053	-0.082
	1.07	0.90	-0.73
AGGILLIQ	0.336	0.084	-0.253
	3.39	1.64	-2.77
AGGTURN	-0.059	-0.028	0.031
	-1.46	-1.69	0.92
RMRF	-0.325	0.837	1.162
	-4.37	34.26	19.90
HML	0.108	0.157	0.049
	0.89	2.80	0.49
SMB	-0.283	0.428	0.710
	-4.69	8.90	7.19
Intercept	0.015	0.007	-0.007
	1.77	2.16	-1.07
N	323	323	323
Adj. R^2	0.384	0.912	0.857

just as in Table IV, our main focus is directed toward the long-short portfolio, presented in the first column of the table. The significantly negative coefficient for mutual fund flows supports that cross-sectional mispricing is exacerbated by mutual fund flows over this time frame as well. These results validate the first part of our hypothesis. Further, by looking at the long and short components separately, we still observe a clear asymmetry between the legs. Mutual fund flows are unrelated to the long leg of the portfolio ($-0.18, t = -0.95$), while there is a significantly negative relation to the short leg. These results suggest that mutual funds primarily increase mispricing by buying overvalued stocks, which is consistent with our previous findings.

When we examine the relationship between hedge fund flows and mispricing, our regressions reveal a noteworthy pattern. As evident from the insignificant coefficients in the second row of Table VII, hedge funds do not appear to attenuate mispricing. This is a striking difference compared to our results for 1994 to 2020. While the explanation for this pattern is not within the scope of our study, we note an interesting relationship between our results and the general performance of the hedge fund industry. Our results indicate that hedge funds, on an aggregate level, fail to identify mispriced stocks, which may be a plausible explanation for the hedge fund industry's poor performance over the last few years.[8] As we have shown that mispricing still prevails, this result raises questions regarding other potential sources of smart money. This is a relevant topic for further research.

C. *Hypothesis 3: Retail Investor Capital Flows Exacerbate Cross-sectional Mispricing of Stocks*

The third part of the empirical analysis focuses on the relationship between retail investor capital flows and cross-sectional mispricing of stocks.
We begin the section by studying pairwise correlations of the key variables used to test the hypothesis, which are presented in Table VIII. The main difference compared to previous correlation tables is the inclusion of retail investor capital flows. Retail investor capital flows and the long-short portfolio display a negative correlation ($\rho = -0.095, p = 0.25$). The sign of the coefficient gives an initial indication that retail investors may exacerbate mispricing.

To examine our third hypothesis – that retail investor capital flows exacerbate mispricing – we now make use of a new regression, where we replace the fund flows with retail investor capital flows. As mutual fund flows are uncorrelated with retail investor capital flows ($\rho = -0.032, p = 0.69$), the risk for bias in the estimation of the coefficient should be limited when mutual fund flows are excluded from the equation.

Table IX presents the results from our regressions with retail investor capital flows as the independent variable. As expected, the coefficient shows that retail investor capital flows are significantly negatively related to the long-short strategy ($-2.56, t = -2.50$). This result suggests that retail investor capital flows exacerbate aggregate mispricing in the cross section, thereby supporting our third hypothesis. Recall that mutual fund flows exacerbate mispricing by buying overvalued stocks, as outlined in Table VII. On a similar note, retail investor capital flows appear to increase mispricing by buying overvalued stocks *and* by selling undervalued stocks. These results can be understood through the negative relationship with the long component of the aggregate portfolio ($-0.94, t = -1.75$), and the positive relationship with the short component ($1.63, t = 2.51$). Thereby, mispricing of overvalued stocks increases during months when retail investors deposit capital to their accounts, while mispricing of undervalued stocks increases when retail investors redeem capital. We conjectured that retail investor capital flows are even dumber than mutual fund flows, as retail investors are less sophisticated than mutual fund managers. These findings corroborate those expectations.

Our results have implications for literature related to market efficiency. While limits-to-arbitrage literature explains why mispricing may persist despite the presence of sophis-

[8] Many hedge funds have underperformed their benchmarks in recent years. For example, the annualized return of the HFRU Hedge Fund Composite USD Index has been as low as 3.1% from 2012 to 2020, underperforming most other equity asset classes.

Table VIII
Pairwise Correlations, 2008-2020

This table shows Pearson pairwise correlation estimates from May 2008 to November 2020 for all variables used to test our third hypothesis. The flow variables are the monthly mean of aggregate equity mutual fund flows (MFFLOW), equity hedge fund flows (HFFLOW) and retail investor capital flows (RIFLOW). LONG and SHORT represent the monthly excess return series of the long and short legs in the mispricing metric (L-S), constructed from the eleven anomalies described in Section III. Control variables are monthly excess market returns (RMRF), aggregate illiquidity (AGGILLIQ), aggregate turnover (AGGTURN), the value factor (HML) and the size factor (SMB). *P*-values are listed below each estimate.

	Pairwise Correlations									
Variable	MFFLOW	HFFLOW	RIFLOW	LONG	SHORT	L-S	RMRF	AGGILLIQ	AGGTURN	HML
HFFLOW	0.129									
	0.11									
RIFLOW	-0.032	-0.238								
	0.69	0.00								
LONG	0.346	0.124	-0.058							
	0.00	0.13	0.48							
SHORT	0.338	0.115	0.006	0.891						
	0.00	0.16	0.94	0.00						
L-S	-0.191	-0.055	-0.095	-0.377	-0.756					
	0.02	0.50	0.25	0.00	0.00					
RMRF	0.336	0.131	-0.073	0.962	0.907	-0.464				
	0.00	0.11	0.37	0.00	0.00	0.00				
AGGILLIQ	-0.223	-0.542	0.215	-0.252	-0.256	0.160	-0.272			
	0.01	0.00	0.01	0.00	0.00	0.05	0.00			
AGGTURN	-0.110	-0.316	-0.008	-0.001	0.053	-0.109	0.023	0.504		
	0.18	0.00	0.92	0.99	0.52	0.18	0.78	0.00		
HML	0.192	0.145	-0.151	0.294	0.320	-0.229	0.325	-0.184	-0.126	
	0.02	0.08	0.06	0.00	0.00	0.00	0.00	0.02	0.12	
SMB	0.142	-0.013	0.157	0.508	0.539	-0.367	0.383	-0.028	0.161	0.202
	0.08	0.87	0.05	0.00	0.00	0.00	0.00	0.73	0.05	0.01

Table IX
Aggregate Retail Investor Capital Flows and Cross-sectional Mispricing, 2008-2020

This table shows the coefficients of time series regressions from May 2008 to December 2020, with the return series of the long-short strategy (L-S), the long leg (LONG) and the short leg (SHORT) as the dependent variables. The portfolios are constructed from the aggregate measure of eleven anomalies. The independent variable is aggregate retail investor capital flows (RIFLOW). Control variables are aggregate illiquidity (AGGILLIQ), aggregate turnover (AGGTURN), the excess return of the market (RMRF), the size factor (SMB) and the value factor (HML).

$$R_{p,t} = \alpha_t + \beta_1 \times RIFLOW_t + \beta_2 \times AGGILLIQ_t + \beta_3 \times AGGTURN_t + \beta_4 \times RMRF_t + \beta_5 \times SMB_t + \beta_6 \times HML_t + \epsilon_{p,t}$$

The t-statistics are listed below the coefficient estimates and are based on Newey-West standard errors.

	Mispricing Metric		
Variable	L-S	LONG	SHORT
RIFLOW	-2.564	-0.939	1.625
	-2.50	-1.75	2.51
AGGILLIQ	0.894	0.325	-0.569
	1.91	2.08	-1.34
AGGTURN	-0.149	-0.099	0.050
	-2.75	-4.30	1.00
RMRF	-0.233	0.935	1.168
	-2.73	43.78	14.98
HML	-0.115	-0.082	0.033
	-0.85	-2.29	0.26
SMB	-0.230	0.370	0.600
	-1.40	8.07	3.40
Intercept	0.037	0.022	-0.015
	3.04	3.91	-1.37
N	152	152	152
Adj. R^2	0.266	0.953	0.863

ticated investors, our results contribute with an explanation of why the pricing anomalies arise in the first place. We conclude that mispricing is likely to prevail as long as *dumber money*, provided by retail investors, enters financial markets. The enduring effect could be explained by a continuing spiral, where *dumb* and *dumber money* increase mispricing, followed by a minority of informed traders, *smart money*, who profit from pricing deviations.

Moreover, our results suggest that increasing stock market participation among retail investors may have real adverse allocation effects in financial markets and that their trading activities are likely to result in value destruction. This is an issue that is highly relevant for policy makers when regulating capital markets to improve market efficiency. Common intuition suggests that the issue is attributable to financial illiteracy among this group of investors, but our results call for further research on the topic.

V. Robustness Checks and Corroborating Evidence

To assess the robustness of our findings, we conduct four robustness checks and tests for corroborating evidence. The first test aims to investigate the robustness of all three hypotheses, while the following tests focus on our inferences regarding retail investor capital flows.

A. *The Momentum Factor*

First, we test if our first conclusion – that mutual fund flows accentuate mispricing while hedge fund flows correct mispricing – is robust when adding the momentum factor to our regressions. Momentum, as discovered by Jegadeesh and Titman (1993), is one of the strongest predictors of stock returns, and the constructed mispricing metric has a significant and strong correlation with the momentum factor. For this reason, we want to rule out the possibility that our results are driven by this anomaly.

First, we assess whether the return predictability of the aggregate mispricing metric persists when controlling for the momentum factor. The four-factor alphas are shown in Panel A of Table X, and we notice that the alpha of the long-short strategy, presented in the first column, is close to identical to the baseline regression. As such, the new regression confirms that the long-short strategy captures mispricing when the momentum factor is included in the regressions. The alphas of the long and short components are significant as well. Further, as seen in Panel B of Table X, all individual anomalies continue to show positive alphas, and the majority of the alphas are significant.[9] Naturally, the alpha of the momentum anomaly is no longer significant (t = 2.89 versus t = 1.59), and the *r-squared* increases notably (0.118 versus 0.877). This change is expected, as the momentum variable is included on both sides of the equation.

When we extend the time frame until 2020, the alphas of the aggregate mispricing metric and the individual anomaly returns are quantitatively similar to the ones presented in Table X. Hence, the mispricing measure from our baseline regressions are robust to including the momentum factor. The tables for the regressions from 1994 to 2020 are presented in Appendix F.

In Table XI, we present results of the regressions used to assess mutual fund flows and hedge fund flows in relation to cross-sectional mispricing, when controlling for the momentum factor. Mutual fund flows still appear to have a significant effect on cross-sectional mispricing ($-1.00, t = -3.07$), but the effect induced by hedge fund flows has diminished markedly, which renders a rather pale link between hedge funds and *smart money*. This vulnerability casts doubt on our previous conclusions related to hedge fund flows, as it appears as if hedge funds primarily trade according to the momentum effect. However, as previously described, there is an ongoing debate regarding which stock or firm characteristics that represent risk or mispricing (Bloomfield and Michaely (2004)). Either way, it is noteworthy that hedge fund flows cannot be defined as *smart money* when classifying the momentum factor as a source of risk, rather than mispricing. When analyzing the effect of

[9] The magnitude of the alphas and the t-values are in line with the ones displayed by Akbas et al. (2015), see Appendix E.

Table X
Fama-French Four-factor Model: Mispricing Metric, 1994-2012

Panel A shows Fama-French four-factor alphas for the cross-sectional investment strategy (L-S) and its two legs (LONG and SHORT) from January 1994 to December 2012. Thereby, control variables are the excess return of the market (RMRF), the size factor (SMB), the value factor (HML) and the momentum factor (UMD):

$$R_{p,t} = \alpha_p + \beta_1 \times RMRF_t + \beta_2 \times SMB_t + \beta_3 \times HML_t + \beta_4 \times UMD_t + \epsilon_{p,t}$$

Panel B shows the intercepts from regressions of eleven individual anomalies that are the basis for the long-short portfolio. The anomalies are return on assets (ROA), Ohlson O-score (OO), failure probability (FP), gross profitability (GP), net stock issues (NSI), total accrual (TA), composite equity issues (CEI), investment-to-assets (ITA), net operating assets (NOA), asset growth (AG) and momentum (M). The same Fama-French four-factor model is used. For brevity, the coefficients for the four factors are not tabulated. The t-statistics are listed below the coefficient estimates and are based on Newey-West standard errors.

Panel A: Fama-French Four-factor Alphas

Variable	L-S	LONG	SHORT
Alpha	0.0145	0.0054	-0.0091
	8.82	5.36	-8.32
RMRF	-0.2667	0.8347	1.1014
	-6.22	29.29	37.89
HML	0.2190	0.2683	0.0494
	3.02	4.62	1.07
SMB	-0.3687	0.4829	0.8517
	-9.12	7.77	17.61
UMD	0.4298	0.0604	-0.3693
	11.92	2.99	-10.31
N	228	228	228
Adj. R^2	0.689	0.919	0.941

Panel B: Individual Anomaly Returns

	ROA	OO	FP	GP	NSI	TA	CEI	ITA	NOA	AG	M
FF4 Alphas	0.0085	0.0066	0.0100	0.0075	0.0073	0.0024	0.0068	0.0034	0.0068	0.0027	0.0032
	3.53	3.30	3.07	3.37	3.97	1.00	3.73	1.68	3.76	1.94	1.59
N	228	228	228	228	228	228	228	228	228	228	228
Adj. R^2	0.540	0.434	0.646	0.428	0.360	0.046	0.450	0.025	0.080	0.396	0.877

Table XI
Fama-French Four-factor Model: Aggregate Mutual Fund Flows, Hedge Fund Flows and Cross-Sectional Mispricing, 1994-2012

This table shows the coefficients of time series regressions with the return series of the long-short strategy (L-S), the long leg (LONG) and the short leg (SHORT) as the dependent variables. The portfolios are constructed from the aggregate measure of eleven anomalies. The independent variables are aggregate mutual fund flows (MFFLOW) and aggregate hedge fund flows (HFFLOW). Control variables are aggregate illiquidity (AGGILLIQ), aggregate turnover (AGGTURN), the excess return of the market (RMRF), the size factor (SMB), the value factor (HML) and the momentum factor (UMD).

$$R_{p,t} = \alpha_t + \beta_1 \times MFFLOW_t + \beta_2 \times HFFLOW_t + \beta_3 \times AGGILLIQ_t + \beta_4 \times AGGTURN_t \\ + \beta_5 \times RMRF_t + \beta_6 \times SMB_t + \beta_7 \times HML_t + \beta_8 \times UMD_t + \epsilon_{p,t}$$

The t-statistics are listed below the coefficient estimates and are based on Newey-West standard errors.

	Mispricing Metric		
Variable	L-S	LONG	SHORT
MFFLOW	-1.002	-0.287	0.715
	-3.07	-1.51	3.19
HFFLOW	0.104	0.040	-0.063
	0.90	0.51	-0.82
AGGILLIQ	0.425	0.126	-0.298
	3.52	2.31	-3.73
AGGTURN	0.048	-0.008	-0.057
	1.13	-0.38	-1.74
RMRF	-0.211	0.846	1.057
	-5.33	28.85	37.45
HML	0.219	0.260	0.041
	4.00	5.11	1.02
SMB	-0.345	0.491	0.836
	-8.09	8.07	17.40
UMD	0.431	0.055	-0.377
	11.92	2.55	-11.88
Intercept	-0.004	0.003	0.007
	-0.47	0.64	1.02
N	228	228	228
Adj. R^2	0.720	0.922	0.946

retail investor capital flows on cross-sectional mispricing while controlling for the momentum factor, the conclusions are essentially unchanged compared to when the momentum factor is excluded. See Appendix F for the tabulated results.

B. Future Returns Predictability

Now, we aim our attention at the inferences about retail investor activity, and conduct our second test. If retail investor capital flows represent *dumber money*, we would expect cross-sectional mispricing to be exacerbated in precisely the same month as capital inflows occur. *Current* retail investor capital flows should therefore be unrelated to *future* returns

Table XII
Aggregate Retail Investor Capital Flows and Future Cross-Sectional Mispricing, 2008-2020

This table shows the coefficients of time series regressions with the future return series of the long-short strategy (L-S), the long leg (LONG) and the short leg (SHORT) as the dependent variables from May 2008 to December 2020. The portfolios are constructed from the aggregate measure of eleven anomalies and are measured over a forward one-month window ($t+1$). The independent variable is aggregate retail investor capital flows (RIFLOW). Control variables are aggregate illiquidity (AGGILLIQ), aggregate turnover (AGGTURN), the excess return of the market (RMRF), the size factor (SMB) and the value factor (HML).

$$R_{p,t} = \alpha_t + \beta_1 \times RIFLOW_t + \beta_2 \times AGGILLIQ_t + \beta_3 \times AGGTURN_t + \beta_4 \times RMRF_t + \beta_5 \times SMB_t + \beta_6 \times HML_t + \beta_7 \times UMD_t + \epsilon_{p,t}$$

The t-statistics are listed below the coefficient estimates and are based on Newey-West standard errors.

	One-month Forward Return (t+1)		
Variable	L-S	LONG	SHORT
RIFLOW	-0.439	0.087	0.527
	-0.45	0.23	0.62
AGGILLIQ	0.711	0.237	-0.474
	1.31	1.61	-1.03
AGGTURN	-0.125	-0.089	0.035
	-2.05	-4.14	0.68
RMRF	-0.228	0.936	1.165
	-2.42	41.63	14.65
HML	-0.083	-0.071	0.012
	-0.60	-2.10	0.10
SMB	-0.291	0.350	0.641
	-2.04	6.83	3.93
Intercept	0.023	0.016	-0.007
	1.58	3.78	-0.59
N	151	151	151
Adj. R^2	0.254	0.951	0.863

of the mispriced stocks. To test this conjecture, we study the relation between capital flows and the subsequent returns of the mispriced stocks. The results are presented in Table XII, where we measure capital flows in month t and portfolio returns in month $t+1$.

The first row of Table XII reveals that, as expected, there is no relationship between retail investor capital flows and the future returns of the mispricing metric ($-0.440, t = -0.44$). If stock prices would have reverted to fair value immediately following months of capital inflows, we would have seen a significantly positive relation. Instead, as previously mentioned, limits to arbitrage are likely to postpone such corrections (Abreu and Brunnermeier (2002)), and it is reasonable to believe that mispricing is attenuated over a longer time window than one month, especially as our previous results show that hedge fund flows do not attenuate mispricing over this period. Altogether, our results suggest that retail investor capital flows have an immediate effect on cross-sectional mispricing, and that the correction of mispricing occurs gradually thereafter. Nevertheless, the evidence of only contemporaneous movements in mispricing and retail investor capital flows corroborates our hypothesis.

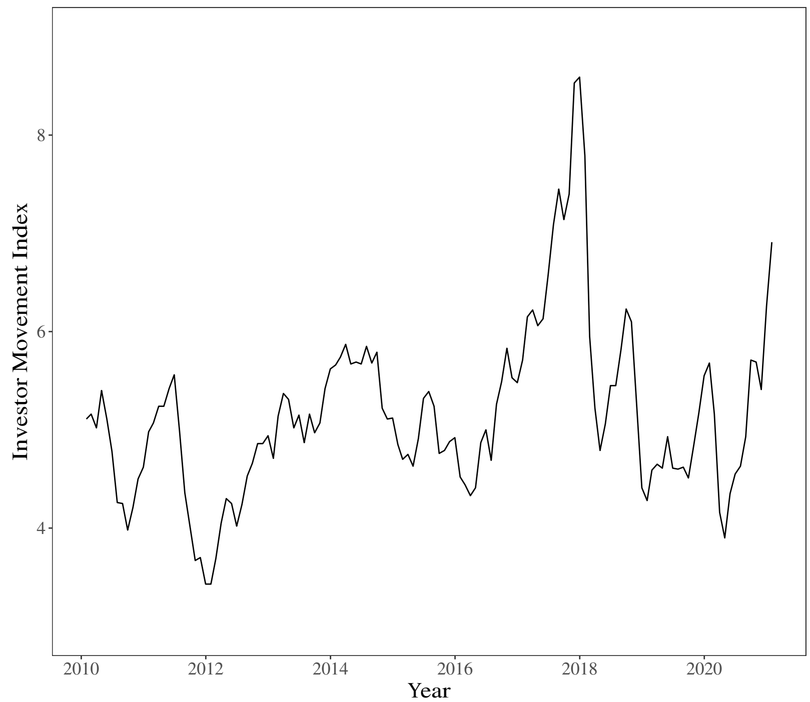

Figure 1. Investor Movement Index, 2010-2020. The Investor Movement Index is used as a proxy for retail investor capital flows and is used to test the robustness of our findings regarding retail investor behavior. The figure illustrates the variation in the Investor Movement Index from January 2010 to December 2020. The index, provided by TD Ameritrade, is created by analyzing the trading activities of the 11 million retail investor accounts of TD Ameritrade.

C. Investor Movement Index

In our third test, we examine our approach to measuring the activity of retail investors. When we use net new assets to retail brokerage houses to measure retail investor capital flows, we do not capture when investors redistribute their already deposited investment base, nor the asset classes held. It is important to rule out potential noise related to these issues. For this reason, we introduce a proxy for retail investor capital flows: the Investor Movement Index.

The index, depicted in Figure 1, reflects the actual trading activities of retail investors by measuring how they are positioned in the market. The Investment Movement Index is provided by one of the largest American brokerage firms, TD Ameritrade. A high absolute value of the metric indicates a bullish sentiment and a low value indicates bearish sentiment. By analyzing a sample from TD Ameritrade's over 11 million funded accounts, the index provides a quantitative measure that captures a self-directed perspective of retail investor capital allocation. Thereby, it is a more detailed measure that circumvents the drawbacks of the data used for retail investor capital flows. The distribution is representative in terms of age, investor experience and account sizes.

There are many indices that measure investor sentiment. For example, Kumar and Lee (2006) create a measure closely related to the Investor Movement Index, as it is used to analyze how retail investors affect stock returns. However, their indicator focuses on

Table XIII
Retail Investor Movement Index and Cross-Sectional Mispricing, 2010-2020

This table shows the coefficients of time series regressions with the return series of the long-short strategy (L-S), the long leg (LONG) and the short leg (SHORT) as the dependent variables from February 2010 to December 2020. The portfolios are constructed from the aggregate measure of eleven anomalies. The independent variable is the relative change in Investor Movement Index (RIMX). Control variables are aggregate illiquidity (AGGILLIQ), aggregate turnover (AGGTURN), the excess return of the market (RMRF), the size factor (SMB) and the value factor (HML).

$$R_{p,t} = \alpha_t + \beta_1 \times RIMX_t + \beta_2 \times AGGILLIQ_t + \beta_3 \times AGGTURN_t + \beta_4 \times RMRF_t + \beta_5 \times SMB_t + \beta_6 \times HML_t + \epsilon_{p,t}$$

The t-statistics are listed below the coefficient estimates and are based on Newey-West standard errors.

	Mispricing Metric		
Variable	L-S	LONG	SHORT
RIMX	-0.047	-0.016	0.032
	-1.57	-0.77	1.38
AGGILLIQ	2.003	0.642	-1.361
	1.72	1.30	-1.78
AGGTURN	-0.077	-0.082	-0.005
	-1.59	-3.64	-0.13
RMRF	-0.125	0.961	1.086
	-1.50	37.37	14.00
HML	0.090	-0.045	-0.135
	0.69	-1.13	-1.06
SMB	-0.341	0.295	0.635
	-2.23	5.07	3.60
Intercept	0.002	0.012	0.009
	0.28	2.56	1.58
N	131	131	131
Adj. R^2	0.187	0.944	0.842

the price impact of specific trades, rather than on aggregate capital flows. The Investor Movement Index is a preferred measure for our study as it aims to capture aggregate trading behavior of retail investors. We are interested in the monthly development of the index and thereby compute the metric used in our regressions as:

$$RIMX_t = \frac{IMX_t}{IMX_{t-1}} - 1, \tag{7}$$

where IMX_t is the Investor Movement Index for month t. Investor Movement Index data is obtained from TD Ameritrade and includes all accessible data, which covers January 2010 until December 2020.[10]

An untabulated analysis of RIMX in relation to our other variables shows that RIMX

[10] Since we compute the relative change in the Investor Movement Index, our regressions include 131 monthly observations from February 2010 until December 2020.

and the long-short strategy are significantly negatively correlated ($\rho = -0.196, p = 0.03$). The sign of the coefficient suggests that our proxy for retail investors capital flows is indeed related to cross-sectional mispricing.

Table XIII shows the results from the regressions using RIMX as the independent variable. The results convey a similar picture as the regressions using retail investor capital flows. The coefficient for RIMX is negative for the long-short portfolio ($-0.047, t = -1.57$) which suggests that aggregate mispricing is exacerbated during months when retail investors become more bullish. Furthermore, there is a positive relation between RIMX and increased mispricing of both overvalued and undervalued stocks ($0.032, t = 1.38$ and $-0.016, t = -0.77$). While the relevant coefficient estimates have the same signs as the regression in Table IX, the results are not significant. It is plausible that the lack of significance is attributable to the relatively small number of monthly observations ($N = 131$). Either way, the results from these regressions do not strongly support our hypothesis, nor do they suggest that our previous conclusions are erroneous, as all coefficients point in the expected direction.

D. The Effect of the Covid-19 Pandemic

In our fourth and final test, we examine whether our results are driven by the Covid-19 pandemic, as the interest in the stock market escalated when many individuals had more time to trade. Looking at the Investor Movement Index in Figure 1, it dropped remarkably in March 2020, when the pandemic reached the Western world, indicating that the pandemic influenced the direction and magnitude of retail investor capital flows. To test if the relationship between mispricing and retail investor capital flows changed during the pandemic, we perform the same type of regressions as in Table IX, with retail investor capital flows as the independent variable, while excluding observations from 2020. This regression shows similar results as before, thereby confirming that our findings are not solely attributable to the development in 2020.

VI. Conclusion

In this study, we analyze the effect of capital flows on cross-sectional mispricing through a replication and extension of Akbas et al. (2015). We confirm that mutual fund flows represent *dumb money* that exacerbate aggregate cross-sectional mispricing by directing capital to overvalued stocks. Moreover, hedge fund flows represent *smart money* that attenuate aggregate mispricing in the cross section. This relationship holds between 1994 and 2012, and thus corroborates the findings by Akbas et al. (2015). However, on a general level, hedge funds appear to have lost its ability to target mispriced stocks in recent years. Finally, our extension unveils that retail investor capital flows represent *dumber money*, as retail investors increase mispricing from two ends: they buy overvalued stocks and sell undervalued stocks. Thereby, retail investors act against rational investing behavior, which results in real adverse allocation effects.

Our research does not aim to explain the underlying reasons for the behavior of individual retail investors, but we shed light on the high relevance of this topic. By targeting the roots of *dumber money*, the economy as a whole could benefit from better capital allocation and improved market efficiency. Furthermore, while we identify a source of *dumber money* that accentuates mispricing, our research does not identify what sources of capital attenuate cross-sectional mispricing today. This is a topic that calls for further investigation.

REFERENCES

Abreu, Dilip, and Markus K Brunnermeier, 2002, Synchronization risk and delayed arbitrage, *Journal of Financial Economics* 66, 341–360.

Akbas, Ferhat, Will Armstrong, Sorin Sorescu, and Avanidhar Subrahmanyam, 2015, Smart money, dumb money and capital market anomalies, *Journal of Financial Economics* 118, 355–382.

Amihud, Yakov, 2002, Illiquidity and stock returns: Cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

Amihud, Yakov, and Haim Mendelson, 1986, Asset pricing and the bid-ask spread, *Journal of Financial Economics* 17, 223–249.

Baker, Malcolm, and Jeffrey Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *Journal of Finance* 61, 1645–1680.

Barinov, Alexander, 2014, Turnover: Liquidity or uncertainty?, *Management Science* 60, 2478–2495.

Bloomfield, Robert, and Roni Michaely, 2004, Risk or mispricing? From the mouths of professionals, *Financial Management* 61–81.

Boehmer, Ekkehart, Charles M Jones, and Xiaoyan Zhang, 2008, Which shorts are informed?, *Journal of Finance* 63, 491–527.

Brennan, Michael J, Tarun Chordia, and Avanidhar Subrahmanyam, 1998, Alternative factor specifications, security characteristics, and the cross-section of expected stock returns, *Journal of Financial Economics* 49, 345–373.

Breusch, Trevor S, and Adrian R Pagan, 1979, A simple test for heteroscedasticity and random coefficient variation, *Econometrica: Journal of the Econometric Society* 1287–1294.

Campbell, John Y, Jens Hilscher, and Jan Szilagyi, 2008, In search of distress risk, *Journal of Finance* 63, 2899–2939.

Cao, Charles, Yong Chen, Bing Liang, and Andrew W Lo, 2013, Can hedge funds time market liquidity?, *Journal of Financial Economics* 109, 493–516.

Chen, Long, Robert Novy-Marx, and Lu Zhang, 2011, An alternative three-factor model, *Available at SSRN 1418117* .

Cooper, Michael J, Huseyin Gulen, and Michael J Schill, 2008, Asset growth and the cross-section of stock returns, *Journal of Finance* 63, 1609–1651.

Coval, Joshua, and Erik Stafford, 2007, Asset fire sales (and purchases) in equity markets, *Journal of Financial Economics* 86, 479–512.

DALBAR, 2015, Dalbar's 21st annual quantitative analysis of investor behavior 2015.

Daniel, Kent, and Sheridan Titman, 2006, Market reactions to tangible and intangible information, *Journal of Finance* 61, 1605–1643.

Datar, Vinay T, Narayan Y Naik, and Robert Radcliffe, 1998, Liquidity and stock returns: An alternative test, *Journal of Financial Markets* 1, 203–219.

Durbin, James, and Geoffrey S Watson, 1950, Testing for serial correlation in least squares regression: I, *Biometrika* 37, 409–428.

Fama, Eugene F, Kenneth R French, David G Booth, and Rex Sinquefield, 1993, Differences in the Risks and Returns of NYSE and NASD Stocks, *Financial Analysts Journal* 49, 37–41.

Frazzini, Andrea, and Owen A Lamont, 2008, Dumb money: Mutual fund flows and the cross-section of stock returns, *Journal of Financial Economics* 88, 299–322.

Fung, William, and David A Hsieh, 2000, Performance characteristics of hedge funds and commodity funds: Natural vs. spurious biases, *Journal of Financial and Quantitative Analysis* 291–307.

Hirshleifer, David, Kewei Hou, Siew Hong Teoh, and Yinglei Zhang, 2004, Do investors overvalue firms with bloated balance sheets?, *Journal of Accounting and Economics* 38, 297–331.

Huang, Jennifer, Clemens Sialm, and Hanjiang Zhang, 2011, Risk shifting and mutual fund performance, *The Review of Financial Studies* 24, 2575–2616.

International Monetary Fund, 2014, Global financial stability report: Moving from liquidity- to growth-driven markets.

Jagannathan, Ravi, Alexey Malakhov, and Dmitry Novikov, 2010, Do Hot Hands Exist Among Hedge Fund Managers? An Empirical Evaluation, *Journal of Finance* 65, 217–255.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Kumar, Alok, and Charles MC Lee, 2006, Retail investor sentiment and return comovements, *Journal of Finance* 61, 2451–2486.

Lintner, John, 1965, Security prices, risk, and maximal gains from diversification, *Journal of Finance* 20, 587–615.

Lou, Dong, 2012, A flow-based explanation for return predictability, *The Review of Financial Studies* 25, 3457–3489.

Loughran, Tim, and Jay R Ritter, 1995, The new issues puzzle, *Journal of Finance* 50, 23–51.

Miller, Edward M, 1977, Risk, uncertainty, and divergence of opinion, *Journal of Finance* 32, 1151–1168.

Novy-Marx, Robert, 2013, The other side of value: The gross profitability premium, *Journal of Financial Economics* 108, 1–28.

Ohlson, James A, 1980, Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research* 109–131.

Reinganum, Marc R, 1990, Market Microstructure and Asset Pricing: An Empirical Investigation of NYSE and NASDAQ Securities, *Journal of Financial Economics* 28, 127–147.

Ritter, Jay R, 1991, The long-run performance of initial public offerings, *Journal of Finance* 46, 3–27.

Securities and Exchange Commission Release No. 34-83063, 2018.

Sharpe, William F, 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425–442.

Shleifer, Andrei, and Robert W Vishny, 1997, The limits of arbitrage, *Journal of Finance* 52, 35–55.

Silber, William L, 1975, Thinness in Capital Markets: The Case of the Tel Aviv Stock Exchange, *Journal of Financial and Quantitative Analysis* 129–142.

Sirri, Erik R, and Peter Tufano, 1998, Costly search and mutual fund flows, *Journal of Finance* 53, 1589–1622.

Sloan, Richard G, 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings?, *Accounting Review* 289–315.

Stambaugh, Robert F, Jianfeng Yu, and Yu Yuan, 2012, The short of it: Investor sentiment and anomalies, *Journal of Financial Economics* 104, 288–302.

Titman, Sheridan, KC John Wei, and Feixue Xie, 2004, Capital investments and stock returns, *Journal of Financial and Quantitative Analysis* 39, 677–700.

Watson, Geoffrey S, and James Durbin, 1951, Exact tests of serial correlation using noncircular statistics, *The Annals of Mathematical Statistics* 446–451.

Appendix A. Anomaly descriptions

In this appendix, we describe how we compute the mispricing score for each stock at the end of month t.

Return on assets. The ratio is constructed using quarterly data,

$$Return\ on\ assets_t = \frac{IBQ_q}{ATQ_{q-1}}, \tag{A1}$$

where IBQ_q is quarterly income before extraordinary items and ATQ_q is total assets for quarter q, which is the most recent quarter for which the reporting date, RDQ, precedes the end of month t.

Ohlson O-score. The Ohlson O-score is calculated as

$$\begin{aligned} O_t = &-0.407SIZE_t + 6.03TLTA_t - 1.43WCTA_t + 0.076CLCA_t - 1.72OENEG_t \\ &- 2.37NITA_t - 1.83FUTL_t + 0.285_INTWO_t - 0.521CHIN_t - 1.32, \end{aligned} \tag{A2}$$

where

$$SIZE_t = log(AT_y) \tag{A3}$$

$$TLTA_t = \frac{DLC_y + DLTT_y}{AT_y} \tag{A4}$$

$$WCTA_t = \frac{ACT_y + LCT_y}{AT_y} \tag{A5}$$

$$CLCA_t = \frac{LCT_y}{ACT_y} \tag{A6}$$

$$OENEG_t = \begin{cases} 1 & \text{if } LT_y > AT_y \\ 0 & \text{otherwise} \end{cases} \tag{A7}$$

$$NITA_t = \frac{NI_y}{AT_y} \tag{A8}$$

$$FUTL_t = \frac{PI_y}{LT_y} \tag{A9}$$

$$INTWO_t = \begin{cases} 1 & \text{if } NI_y < 0 \text{ and } NI_{y-1} < 0 \\ 0 & \text{otherwise} \end{cases} \tag{A10}$$

$$CHIN_t = \frac{NI_y - NI_{y-1}}{|NI_y| + |NI_{y-1}|} \tag{A11}$$

and AT_y is total assets, DLC_y is debt in current liabilities, $DLTT_y$ is long-term debt, ACT_y is current assets, LCT_y is current liabilities, LT_y is total liabilities, NI_y is net income and PI_y is pretax income for year y, which is the most recent reporting year that ends at least four months before the end of month t. We do this to direct the metric to the month where the information is assumed to be publicly available, in line with Stambaugh, Yu and Yuan (2012) and other accounting literature (for example Hirshleifer, Hou, Teoh, and Zhang (2004)). A logistic function can be used to obtain the financial distress probability, where P increases in O:

$$P_t = \frac{1}{1 + exp(-O_t)} \tag{A12}$$

Failure probability. Campbell, Hilscher and Szilagyi (2008) use logit regressions to calculate the failure probability. In the following regression, the coefficients are extracted from their 12-month logit regressions:

$$\begin{aligned} x_t = & -20.26NIMTAAVG_t + 1.42TLMTA_t - 7.13EXRETAVG_t + 1.41SIGMA_t \\ & - 0.045RSIZE_t - 2.13CASHMTA_t + 0.075MB_t - 0.058PRICE_t - 9.16, \end{aligned} \tag{A13}$$

where

$$NIMTA_t = \frac{NIQ_q}{LTQ_q + ME_t} \tag{A14}$$

$$TLMTA_t = \frac{LTQ_q}{LTQ_q + ME_t} \tag{A15}$$

$$EXRET_t = log(1 + RET_t) - log(1 + SPRTRN_t) \tag{A16}$$

$$SIGMA_t = \left(252 \times \frac{1}{N-1} \times \sum_{k \in (t, t-1, t-2)} r_k^2\right)^{1/2} \tag{A17}$$

$$RSIZE_t = log\left(\frac{ME_t}{TOTVAL_t}\right) \tag{A18}$$

$$CASHMTA_t = \frac{CHEQ_q}{LTQ_q + ME_t} \tag{A19}$$

$$MB_t = \frac{ME_t}{CEQQ_q + (ME_t - CEQQ_q) \times 10\%} \tag{A20}$$

$$PRICE_t = log(PRC_t) \tag{A21}$$

and $NIMTAAVG_t$ and $EXRETAVG_t$ are defined as

$$NIMTAAVG_t = \frac{1-\phi^3}{1-\phi^{12}}(NIMTA_{t,t-2} + ... + \phi^9 NIMTA_{t-9,t-11}) \tag{A22}$$

$$EXRETAVG_t = \frac{1-\phi}{1-\phi^{12}}(EXRET_t + ... + \phi^{11} EXRET_{t-11}). \tag{A23}$$

NIQ_q is net income and LTQ_q is total liabilities for quarter q. ME_t is market equity capitalization for month t. Missing values for $NIMTA_t$ and $EXRET_t$ are replaced by the respective cross-sectional means. RET_t is the monthly stock return, and $SPRTRN_t$ is the monthly return of the S&P500 index for month t. $SIGMA_t$ is computed using daily returns, r, and at least five non-zero daily observations are required during the past three months k for the metric to be included. $TOTVAL_t$ is the market capitalization of the S&P500 index for month t. $CHEQ_q$ is cash and short-term investments. The constant ϕ equals $2^{-1/3}$ and is included together with a multiplier to assign more recent values a greater weight.

Further, MB_t represents the market-to-book ratio, where the book value is increased by 10% of the difference between the market value and the book value. If the retrieved book value is negative, a book value of USD 1 is used instead. PRC_t is the share price for month t and is truncated at 15 so that $PRICE_t$ can take a maximum value of $log(15)$. All variables except $PRICE_t$ are winsorized at the 5% and 95% level in the cross section. Equity market variables are obtained for month t. For quarterly accounting data, the reporting quarter q is the most recent quarter for which the reporting date, RDQ, precedes the end of month t.

To compute the probability of failure, the obtained value can be inserted in a probability function. A higher value of x_t implies a higher risk of financial distress.

$$P_t = \frac{1}{1+exp(-x_t)} \tag{A24}$$

Gross profitability. Gross profitability is measured as

$$Gross\ profitability_t = \frac{REVT_y - COGS_y}{AT_y}, \tag{A25}$$

where $REVT_y$ is annual total revenue, $COGS_y$ is annual cost of goods sold and AT_y is total assets for year y. Year y the most recent reporting year that ends at least four months before the end of month t.

Net stock issues. This effect is quantified as the annual log change in split-adjusted shares outstanding,

$$Net\ stock\ issues_t = log\left(\frac{CSHO_y \times ADJEX\_C_y}{CSHO_{y-1} \times ADJEX\_C_{y-1}}\right), \tag{A26}$$

where $CSHO_y$ is common shares outstanding and $ADJEX\_C_y$ is the cumulative adjustment factor for year y, which is the reporting year that ends at least four months before the end of month t.

Total accrual. We compute total accrual as the annual change in non-cash working capital minus depreciation and amortization, divided by average total assets for the two previous fiscal years,

$$Total\ accrual_t = \frac{(\Delta ACT_y - \Delta CHE_y - \Delta DLC_y + \Delta LCT_y + \Delta TXP_y) - DP_y}{\frac{AT_y + AT_{y-1}}{2}}, \tag{A27}$$

where ACT_y is total current assets, CHE_y is cash and short-term investments, DLC_y is debt in current liabilities, LCT_y is current liabilities, TXP_y is income taxes payable, DP_y is depreciation and amortization and AT_y is total assets for year y, which is the reporting year that ends at least four months before the end of month t.

Composite equity issues. The predictive value of the anomaly is quantified as the difference between the growth in market capitalization and the 12-month cumulative stock return,

$$Composite\ equity\ issues_t = \left(\frac{ME_{t-4}}{ME_{t-16}} - 1\right) - ((1 + RET_{t-4} \times \ldots \times (1 + RET_{t-15}) - 1), \tag{A28}$$

where ME_t is the market capitalization and RET_t is the monthly stock return for month t. The variables are lagged four months to match the net stock issues anomaly.

Investment-to-assets. Investment-to-assets is computed as

$$Investment\text{–}to\text{–}assets_t = \frac{\Delta PPEGT_y + \Delta INVT_y}{AT_{y-1}}, \tag{A29}$$

where $PPEGT_y$ is gross property, plant and equipment, $INVT_y$ is inventory and AT_y is total assets for year y, which is the reporting year that ends at least four months before the end of month t.

Net operating assets. The measure represents the scaled difference between operating income and free cash flow. It is defined as the difference between operating assets and operating liabilities, divided by total assets,

$$Net\ operating\ assets_t = \frac{(AT_y - CHE_y) - (AT_y - DLC_y - DLTT_y - CEQ_y - MIB_y - PSTK_y)}{AT_{y-1}}, \tag{A30}$$

where AT_y is total assets, CHE_y is cash and short-term investments, DLC_y is debt in current liabilities, $DLTT_y$ is long-term debt, CEQ_y is common equity, MIB_y is minority interest and $PSTK_y$ is preferred stocks for year y, which is the reporting year that ends at least four months before the end of month t. Minority interest and preferred stocks are assigned a value of zero if missing.

Asset growth. Asset growth is measured as the year-over-year growth in total assets,

$$Asset\ growth_t = \frac{AT_y}{AT_{y-1}} - 1, \tag{A31}$$

where AT_y is total assets for year y, which is the reporting year that ends at least four months before the end of month t.

Momentum. To measure the momentum effect, the cumulative return for month $t-11$ to month $t-1$ is computed,

$$Momentum_t = (1 + RET_{t-1}) \times \ldots \times (1 + RET_{t-11}) - 1, \tag{A32}$$

where RET_t is the monthly stock return for month t.

Appendix B. Data Collection

This appendix describes how we filter our mutual fund data. Mutual funds with the following Lipper objectives have been included in our sample: CA, CG, CS, El, FS, G, GI, H, ID, LCCE, LCGE, LCVE, MC, MCCE, MCGE, MCVE, MLCE, MLGE, MLVE, MR, NR, S, SCCE, SCGE, SCVE, SG, SP, TK, TL, UT. We also include funds with a Strategic Insights objective of AGG, ENV, FIN, GMC, GRI, GRO, HLT, ING, NTR, SCG, SEC, TEC, UTI, GLD or RLE as well as funds with a Wiesenberger Fund Type Code of G, G-I, G-S, GCI, IEQ, ENR, FIN, GRI, HLT, LTG, MCG, SCG, TCH, UTL or GPM. Further, we include mutual funds that hold more than 80% of its value in common stock or with a so-called CS policy.

Appendix C. Market Illiquidity

This appendix explains how we compute our control variables for market illiquidity.

Aggregate illiquidity. The monthly illiquidity metric for stock i in month t is computed as

$$ILLIQ_{i,t} = \frac{\sum_{i=1}^{N_{i,t}} \left[\frac{|RET_{i,d,t}|}{VOL_{i,d,t}} \right]}{N_{i,t}}, \tag{C1}$$

where $r_{i,d,t}$ is the daily return for stock i on day d in month t. $VOL_{i,d,t}$ is the daily trading volume in dollars for stock i on day d in month t and $N_{i,t}$ is the number of daily observations for stock i in month t. Thus, the ratio yields the absolute percentage price change per dollar trading volume. Further, the aggregate monthly illiquidity measure is computed as

$$AGGILLIQ_t = \frac{\sum_{i=1}^{N_t} [ILLIQ_{i,t}]}{N_t} \tag{C2}$$

where N_t is the number of observations in month t.

Aggregate turnover. Aggregate turnover is computed as

$$AGGTURN_t = \frac{\sum_{i=1}^{N} \left[\frac{Avg.\ trading\ volume_{i,t}}{SHROUT_{i,t}} \right]}{N_t}, \tag{C3}$$

where $Avg.\ trading\ volume_{i,t}$ is the average trading volume for the most recent three months for stock i and $SHROUT_{i,t}$ is the number of shares outstanding for stock i for month t. N_t is the number of observations in month t.

Appendix D. Empirical Analysis

This appendix presents tables referred to in our empirical analysis. The tables are shown on the following pages.

Table D1
Sharpe Ratios of Long-short Portfolios based on Cross-sectional Return Predictors, 1994-2012

This table presents the Sharpe ratios for the individual portfolio returns of eleven anomalies and the aggregate long-short portfolio (L-S). The anomalies are return on assets (ROA), Ohlson O-score (OO), failure probability (FP), gross profitability (GP), net stock issues (NSI), total accrual (TA), composite equity issues (CEI), investment-to-assets (ITA), net operating assets (NOA), asset growth (AG) and momentum (M). In untabulated tests, we confirm that the Sharpe ratio of the aggregate long-short portfolio is significantly higher than that of each of the individual anomalies in ten out of eleven tests. The insignificant difference is present between the net operating assets anomaly and the aggregate long-short portfolio.

	ROA	OO	FP	GP	NSI	TA	CEI	ITA	NOA	AG	M	L-S
Sharpe ratio	0.506	0.356	0.378	0.409	0.657	0.239	0.582	0.420	0.970	0.382	0.404	1.310

Table D2
Aggregate Mutual Fund Flows, Hedge Fund Flows and Individual Anomaly Returns, 1994-2020

This table shows the coefficients of time series regressions with the return of the long-short strategy (L-S) for eleven individual anomalies as the dependent variable between January 1994 and November 2020. The anomalies are return on assets (ROA), Ohlson O-score (OO), failure probability (FP), gross profitability (GP), net stock issues (NSI), total accrual (TA), composite equity issues (CEI), investment-to-assets (ITA), net operating assets (NOA), asset growth (AG) and momentum (M). The independent variables are aggregate mutual fund flows (MFFLOW) and aggregate hedge fund flows (HFFLOW). Control variables are aggregate illiquidity (AGGILLIQ), aggregate turnover (AGGTURN), the excess return of the market (RMRF), the size factor (SMB) and the value factor (HML).

$$R_{p,t} \;=\; \alpha_t + \beta_1 \times MFFLOW_t + \beta_2 \times HFFLOW_t + \beta_3 \times AGGILLIQ_t + \beta_4 \times AGGTURN_t + \beta_5 \times RMRF_t + \beta_6 \times SMB_t + \beta_7 \times HML_t + \epsilon_{p,t}$$

The t-statistics are listed below the coefficient estimates and are based on Newey-West standard errors.

	ROA	OO	FP	GP	NSI	TA	CEI	ITA	NOA	AG	M
MFFLOW	-0.012	-0.463	0.630	-0.028	-0.358	-0.461	-0.030	-0.024	-0.629	-0.237	-0.019
	-0.02	-1.17	0.50	-0.05	-1.02	-1.05	-0.10	-0.05	-1.89	-0.58	-0.02
HFFLOW	0.091	0.038	0.147	0.006	-0.014	0.040	-0.001	-0.066	-0.024	-0.059	0.434
	0.69	0.30	0.64	0.04	-0.17	0.36	-0.01	-0.49	-0.18	-0.55	1.08
AGGILLIQ	0.119	0.235	-0.177	0.066	0.148	0.167	0.181	0.081	0.237	0.179	-0.333
	0.85	2.56	-0.70	0.50	1.52	1.42	1.71	0.77	3.20	1.81	-1.65
AGGTURN	0.021	-0.034	-0.061	-0.018	0.008	-0.029	0.030	-0.044	-0.017	0.025	-0.202
	0.47	-0.95	-0.96	-0.37	0.23	-0.75	0.75	-0.99	-0.52	0.80	-1.85
RMRF	-0.480	-0.063	-1.107	-0.407	-0.248	0.115	-0.238	-0.113	0.086	-0.204	-0.586
	-7.08	-1.22	-9.02	-5.80	-5.59	2.05	-4.59	-1.86	2.04	-3.05	-4.21
HML	-0.054	-0.329	-0.503	-0.568	0.128	0.153	0.359	0.196	0.151	0.590	-0.831
	-0.56	-5.18	-1.98	-3.92	1.72	1.63	3.79	2.55	1.69	4.97	-2.92
SMB	-0.668	-0.696	-0.497	-0.156	-0.252	-0.148	-0.316	0.102	0.016	0.167	0.030
	-9.35	-11.51	-2.36	-2.27	-4.63	-2.06	-4.90	1.22	0.17	1.49	0.13
Intercept	0.003	0.003	0.027	0.008	0.003	-0.001	-0.004	0.007	0.005	-0.004	0.051
	0.33	0.42	2.07	0.84	0.38	-0.08	-0.44	0.79	0.85	-0.59	2.58
N	323	323	323	323	323	323	323	323	323	323	323
Adj. R^2	0.469	0.389	0.448	0.294	0.324	0.052	0.394	0.064	0.043	0.324	0.187

Appendix E. Original Tables from Akbas et al. (2015)

This appendix shows the original tables from Akbas et al. (2015) to enable a comparison between the results.

Table E4 shows that our results are very similar to those of Akbas et al. (2015), as we obtain positive alphas for all long-short portfolios on both an individual and aggregate level. However, there is a notable difference in the significance of the momentum factor for the aggregate long-short portfolio. While our momenum factor is significant, Akbas et al. (2015) report a t-statistic of 1.50. We find it reasonable that the momentum factor is significant as the momentum anomaly is included in the long-short portfolio, present on the left-hand side of the equation.

On a similar note, when adding the momentum factor to the regressions, we obtain a higher $r\text{–}squared$ for the returns of the individual anomaly momentum compared to Akbas et al. (2015), which can be seen in Panel B of Table E4. We are surprised that Akbas et al. (2015) present such a low $r\text{–}squared$, as the momentum factor should explain a significant share of the variation in the returns of the momentum anomaly.

We generally obtained higher $r\text{–}squared$ for the individual anomalies compared to Akbas et al. (2015), as seen in Panel B of Table E4. The difference is likely due to the individual anomalies being noisy, which Akbas et al. (2015) highlight as well. As the portfolios of the individual anomalies include a limited sample of stocks, they become highly dependent on which particular stocks are included. Further, Akbas et al. (2015) does not provide information about how they construct the portfolios. We follow the closest related paper, Stambaugh, Yu and Yuan (2012), and replicate their method precisely. However, the sole purpose of constructing the anomalies is to capture cross-sectional mispricing, which is fulfilled as most of the alphas are significantly positive. Thus, even though the values of the $r\text{–}squared$ differ between our results and those of Akbas et al. (2015), we can effectively use the portfolios to measure cross-sectional mispricing.

When looking at the relationship between aggregate mutual fund flows and hedge fund flows and the individual anomaly returns in Table E5, our results corroborate those of Akbas et al. (2015). Despite the noisy return series of the individual anomalies, our findings reveal the same pattern, as there is a negative relationship between mutual fund flows and long-short portfolio returns, and mixed and insignificant signs for the coefficients of the hedge fund flows. Akbas et al. (2015) use the analysis in Table E5 as guidance on how to construct suitable mispricing metrics in their following analyses, and our obtained results fulfil this purpose as well.

Table E3
Akbas et al (2015): Summary Statistics 1994-2012

This is the original table from Akbas et al. (2015), which displays univariate summary statistics for all variables between January 1994 and December 2012. The flow variables are the monthly mean of aggregate equity mutual fund flows (MFFLOW) and equity hedge fund flows (HFFLOW). LONG and SHORT represent the monthly excess return series of the long and the short legs in the mispricing metric L-S, constructed from the eleven anomalies described in Section III. Control variables are the monthly excess return of the market (RMRF), aggregate illiquidity (AGGILLIQ), aggregate turnover (AGGTURN), the value factor (HML) and the size factor (SMB). The correlation table in Panel B shows the Pearson pairwise correlation estimates. P-values are listed below each estimate.

Panel A: Descriptive statistics, 1994–2012

Variable	N	Mean	Median	Standard deviation	Minimum	10th percentile	25th percentile	75th percentile	90th percentile	Maximum
MFFLOW	228	0.0033	0.0030	0.005	−0.015	−0.003	0.000	0.006	0.010	0.022
HFFLOW	228	0.0076	0.0095	0.019	−0.102	−0.011	0.000	0.019	0.025	0.074
LONG	228	0.0138	0.0142	0.053	−0.189	−0.055	−0.015	0.050	0.074	0.176
SHORT	228	−0.0060	0.0015	0.074	−0.273	−0.100	−0.053	0.042	0.085	0.185
L−S	228	0.0198	0.0171	0.040	−0.102	−0.023	−0.003	0.037	0.068	0.157
RMRF	228	0.0052	0.0119	0.046	−0.172	−0.058	−0.022	0.035	0.061	0.113
AGGILLIQ	228	0.0427	0.0405	0.024	0.008	0.016	0.022	0.058	0.078	0.128
AGGTURN	228	0.1460	0.1242	0.076	0.049	0.067	0.080	0.199	0.258	0.403
HML	228	0.0045	0.0025	0.033	−0.098	−0.031	−0.014	0.019	0.042	0.138
SMB	228	−0.0012	−0.0016	0.035	−0.220	−0.037	−0.021	0.022	0.037	0.077

Panel B: Pairwise correlations, 1994–2012

Variable	MFFLOW	HFFLOW	LONG	SHORT	L-S	RMRF	AGGILLIQ	AGGTURN	HML
HFFLOW	0.173								
	0.01								
LONG	0.309	0.072							
	0.00	0.28							
SHORT	0.318	−0.015	0.854						
	0.00	0.82	0.00						
L-S	−0.177	0.125	−0.247	−0.715					
	0.01	0.06	0.00	0.00					
RMRF	0.345	0.040	0.836	0.855	−0.470				
	0.00	0.55	0.00	0.00	0.00				
AGGILLIQ	0.479	−0.171	0.018	−0.056	0.130	0.032			
	0.00	0.01	0.78	0.40	0.05	0.63			
AGGTURN	−0.544	−0.179	−0.134	−0.012	−0.158	−0.117	−0.602		
	0.00	0.01	0.04	0.86	0.02	0.08	0.00		
HML	0.024	0.185	−0.119	−0.259	0.323	−0.204	−0.001	−0.104	
	0.71	0.01	0.07	0.00	0.00	0.00	0.99	0.12	
SMB	0.037	−0.029	0.392	0.433	−0.280	0.190	−0.145	0.080	−0.353
	0.58	0.67	0.00	0.00	0.00	0.00	0.03	0.23	0.00

Table E4
Akbas et al. (2015): Mispricing Metric, 1994-2012

This is the original table from Akbas et al. (2015). The left-hand side of Panel A highlights the mean excess returns of a long-short portfolio (L-S), which is a cross-sectional strategy constructed to capture mispricing, and the long and short components (LONG and SHORT). The right-hand side of Panel A shows Fama-French three-factor alphas for the cross-sectional investment strategy (L-S) and its two legs (LONG and SHORT) in the regression:

$$R_{p,t} = \alpha_p + \beta_1 \times RMRF_t + \beta_2 \times SMB_t + \beta_3 \times HML_t + \epsilon_{p,t}$$

Panel B shows the intercepts from regressions of eleven individual anomalies that are the basis for the long-short portfolio. The same Fama-French three-factor model is used. The t-statistics are listed below the coefficient estimates and are based on Newey-West standard errors.

Panel A: Mispricing metric returns

	Mean excess returns			Fama and French three-factor alphas			Fama and French four-factor alphas		
Variable	L−S	LONG	SHORT	L−S	LONG	SHORT	L−S	LONG	SHORT
Alpha	0.0198	0.0138	−0.0060	0.0204	0.0084	−0.0120	0.0187	0.0060	−0.0127
	6.26	3.88	−1.15	8.06	4.73	−5.20	8.28	3.49	−5.94
RMRF				−0.3506	0.9429	1.2935	−0.3471	0.9478	1.2949
				−5.48	14.30	19.89	−5.63	14.88	19.50
HML				0.2312	0.2409	0.0097	0.2217	0.2276	0.0059
				1.64	2.90	0.07	1.63	2.77	0.04
SMB				−0.1515	0.4402	0.5916	−0.1564	0.4333	0.5897
				−2.27	2.67	3.01	−2.33	2.81	3.00
UMD							0.1191	0.1662	0.0472
							1.50	1.97	0.61
N	228	228	228	228	228	228	228	228	228
Adj. R^2				0.281	0.771	0.805	0.303	0.797	0.805

Panel B: Individual anomaly returns

	Return on assets (1)	Ohlson O-score (2)	Failure probability (3)	Gross profitability (4)	Net stock issues (5)	Total accrual (6)	Composite equity issues (7)	Investment- to-assets (8)	Net operating assets (9)	Asset growth (10)	Momentum (11)
FF3 alphas	0.0060	0.0066	0.0227	0.0094	0.0105	0.0042	0.0143	0.0035	0.0098	0.0089	0.0163
	1.82	2.88	8.97	3.14	4.78	2.53	4.69	2.33	2.75	3.61	4.37
Adj. R^2	0.110	0.054	0.241	0.079	0.386	0.021	0.081	0.035	0.092	0.173	0.050
FF4 alphas	0.0092	0.0094	0.0249	0.0116	0.0115	0.0030	0.0092	0.0034	0.0069	0.0064	0.0104
	2.45	2.92	9.43	3.72	5.04	1.89	3.18	2.12	1.86	2.60	2.08
Adj. R^2	0.151	0.109	0.268	0.097	0.391	0.052	0.233	0.031	0.133	0.234	0.135

Table E5

Akbas et al (2015): Aggregate Mutual Fund Flows, Hedge Fund Flows and Individual Anomaly Returns, 1994-2012

This is the original table from Akbas et al. (2015), and shows the coefficients of time series regressions with the return of the long-short strategy (L-S) for eleven individual anomalies as the dependent variable. The independent variables are aggregate mutual fund flows (MFFLOW) and aggregate hedge fund flows (HFFLOW). Control variables are aggregate illiquidity (AGGILLIQ), aggregate turnover (AGGTURN), the excess return of the market (RMRF), the size factor (SMB) and the value factor (HML).

$$R_{p,t} = \alpha_p + \beta_1 \times MFFLOW_t + \beta_2 \times HFFLOW_t + \beta_3 \times AGGILLIQ_t + \beta_4 \times AGGTURN_t + \beta_5 \times RMRF_t + \beta_6 \times SMB_t + \beta_7 \times HML_t + \epsilon_{p,t}$$

The t-statistics are listed below the coefficient estimates and are based on Newey-West standard errors.

Individual anomaly returns											
Variable	Return on assets	Ohlson O-score	Failure probability	Gross profitability	Net stock issues	Total accruals	Composite equity issues	Investment-to-assets	Net operating assets	Asset growth	Momentum
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
MFFLOW	−1.840	−1.755	−2.111	−1.833	−1.311	−0.988	−0.192	0.546	0.764	0.188	−1.846
	−1.91	−2.53	−2.90	−2.00	−2.11	−2.88	−0.26	1.58	0.88	0.47	−1.91
HFFLOW	0.151	0.216	0.147	0.226	0.010	0.103	−0.074	0.077	0.012	−0.066	0.295
	0.91	1.31	1.11	2.03	0.09	1.54	−0.45	0.86	0.07	−0.54	0.87
RMRF	−0.335	−0.184	−0.439	−0.185	−0.386	0.079	0.344	−0.138	0.106	−0.130	−0.194
	−3.76	−2.67	−5.33	−2.13	−5.54	1.52	3.45	−2.75	1.06	−1.83	−1.07
AGGILLIQ	0.303	0.477	−0.048	0.267	0.229	0.097	0.284	0.074	0.136	0.027	−0.309
	2.19	4.88	−0.39	2.08	2.45	1.47	1.88	0.88	0.88	0.32	−1.36
AGGTURN	−0.016	0.032	−0.152	−0.034	−0.018	−0.053	0.026	0.027	−0.003	−0.048	−0.287
	−0.35	0.99	−3.17	−0.89	−0.55	−2.30	0.41	0.69	−0.04	−1.46	−3.68
HML	0.182	−0.026	−0.052	0.258	0.484	0.038	0.081	−0.056	−0.403	0.391	0.007
	1.43	−0.33	−0.36	1.96	4.92	0.61	0.56	−0.93	−3.23	3.76	0.03
SMB	0.062	0.032	−0.139	0.380	−0.012	−0.081	−0.014	0.037	−0.374	−0.037	−0.360
	0.25	0.20	−1.04	2.00	−0.09	−1.01	−0.09	0.91	−1.66	−0.35	−1.82
Intercept	0.000	−0.015	0.053	0.007	0.007	0.010	−0.001	−0.006	0.002	0.015	0.076
	−0.01	−1.73	4.47	0.68	0.87	1.64	−0.04	−0.69	0.13	1.63	3.68
N	228	228	228	228	228	228	228	228	228	228	228
Adj. R^2	0.114	0.085	0.281	0.086	0.394	0.049	0.080	0.036	0.091	0.175	0.096

Appendix F. Robustness Check and Corroborative Evidence

This appendix shows tables presenting results related to our robustness tests. The tables are presented on the following pages.

Table F6

Fama-French Four-factor Model: Mispricing Metric, 1994-2020

Panel A shows Fama-French four-factor alphas for the cross-sectional investment strategy (L-S) and its two legs (LONG and SHORT) in the regression:

$$R_{p,t} = \alpha_p + \beta_1 \times RMRF_t + \beta_2 \times SMB_t + \beta_3 \times HML_t + \beta_4 \times UMD_t + \epsilon_{p,t}$$

Panel B shows the intercepts from regressions of eleven individual anomalies that are the basis for the long-short portfolio. The anomalies are return on assets (ROA), Ohlson O-score (OO), failure probability (FP), gross profitability (GP), net stock issues (NSI), total accrual (TA), composite equity issues (CEI), investment-to-assets (ITA), net operating assets (NOA), asset growth (AG) and momentum (M). The same Fama-French four-factor model is utilized. For brevity, the coefficients for the four factors are not tabulated. The t-statistics are listed below the coefficient estimates and are based on Newey-West standard errors.

Panel A: Fama-French Four-factor Alphas

Variable	L-S	LONG	SHORT
Alpha	0.0106	0.0047	-0.0060
	6.20	4.81	-4.16
RMRF	-0.2245	0.8549	1.0794
	-4.63	37.18	27.24
HML	0.3092	0.2030	-0.1062
	4.28	3.78	-1.62
SMB	-0.3234	0.4206	0.7440
	-7.28	7.61	11.89
UMD	0.4069	0.0750	-0.3319
	11.17	3.57	-9.34
N	323	323	323
Adj. R^2	0.564	0.914	0.898

Panel B: Individual Anomaly Returns

	ROA	OO	FP	GP	NSI	TA	CEI	ITA	NOA	AG	M
FF4 Alphas	0.0078	0.0037	0.0098	0.0056	0.0069	-0.0007	0.0056	0.0023	0.0074	0.0035	0.0036
	3.87	2.15	3.20	2.51	4.63	-0.38	3.48	1.25	4.59	2.23	2.06
N	323	323	323	323	323	323	323	323	323	323	323
Adj. R^2	0.533	0.369	0.633	0.339	0.331	0.050	0.393	0.057	0.053	0.322	0.822

Table F7
Fama-French Four-Factor Model: Aggregate Retail Investor Capital Flows and Cross-sectional Mispricing, 2008-2020

This table shows the coefficients of time series regressions with the return series of the long-short strategy (L-S), the long leg (LONG) and the short leg (SHORT) as the dependent variables. The portfolios are constructed from the aggregate measure of eleven anomalies. The independent variable is retail investor capital flows (RIFLOW). Control variables are aggregate illiquidity (AGGILLIQ), aggregate turnover (AGGTURN), the excess return of the market (RMRF), the size factor (SMB), the value factor (HML) and the momentum factor (UMD).

$$R_{p,t} = \alpha_t + \beta_1 \times RIFLOW_t + \beta_2 \times AGGILLIQ_t + \beta_3 \times AGGTURN_t + \beta_4 \times RMRF_t + \beta_5 \times SMB_t + \beta_6 \times HML_t + \beta_7 \times UMD_t + \epsilon_{p,t}$$

The t-statistics are listed below the coefficient estimates and are based on Newey-West standard errors.

	Mispricing Metric		
	L-S	LONG	SHORT
RIFLOW	-2.700	-0.961	1.739
	-2.82	-1.73	2.97
AGGILLIQ	1.638	0.446	-1.192
	6.58	2.26	-5.19
AGGTURN	-0.123	-0.095	0.028
	-3.96	-4.42	0.85
RMRF	-0.109	0.955	1.065
	-1.84	42.43	17.28
HML	0.115	-0.044	-0.159
	1.01	-1.25	-1.58
SMB	-0.213	0.373	0.586
	-1.27	7.92	3.33
UMD	0.344	0.056	-0.288
	6.09	2.49	-4.56
Intercept	0.025	0.020	-0.005
	2.80	3.70	-0.64
N	152	152	152
Adj. R^2	0.416	0.954	0.888