The London School of Economics and Political Science

Essays in Corporate Finance and Asset Pricing

Mark Morley

Thesis submitted to the Department of Finance of the London School of Economics and Political Science for the degree of Doctor of Philosophy

December 4, 2025

Acknowledgements

I am deeply indebted to my advisors, Martin Oehmke and Christopher Polk, as well as my third letter writer, Christian Julliard. Without their advice, guidance, and encouragement, writing this thesis would not have been possible.

In addition, I would like to thank the faculty at the London School of Economics for the many valuable comments, questions, suggestions, and discussions, both within and outside the seminar room. I am especially grateful to Michela Verardo, Amil Dasgupta, Mike Burkart, Daniel Ferriera, Dirk Jenter, Juanita Gonzalez-Uribe, Kim Fe Cramer, Peter Kondor, and Ian Martin.

I am grateful to the faculty and staff at the Department of Finance who supported me at the various stages of my PhD journey, especially Mary Comben, Osmana Raie, Simon Tuck, and Miranda Imperial. I acknowledge the financial support of the London School of Economics.

I also want to thank my fellow PhD colleagues who have impacted my time and work substantially: Christopher Grenier, Nicolas Garrido Sureda, Marc Gischer, Magnus Hjortfors Irie, Ali Choubdaran Varnosfaderani, Zheyuan Yang, and Jiahong Shi.

Finally, I am immensely grateful for the support I have received from my parents, friends, and my girlfriend, Yue.

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgment is made. This thesis may not be reproduced without my prior written consent.

I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 45,669 words.

Mark Morley

Abstract

This thesis contains three chapters in asset pricing and corporate finance.

In the first chapter I investigate the hypothesis that (i) poor quality mutual funds gamble to increase their volatility and (ii) they can be found in both tails of the return distribution. Sorting funds into buckets based on their return realizations, I find evidence that poor quality funds sometimes earn extremely high returns over short horizons. These funds have lower excess returns, Sharpe Ratio, and alphas than comparable funds in one to five year horizons. I find that bad funds with high short-term returns follow extreme and transient strategies with high average factor exposures and high turnover.

In the second chapter I investigate the optimal re-hiring rules of investors in mutual funds. Good and bad fund managers are compensated over multiple periods and choose behavior in the presence of career concerns: their performance is evaluated by investors making re-hiring decisions. If investors cannot commit to a re-hiring rule with profitable deviations, they only re-hire a manager who trades exactly like a good manager, forcing the bad manager to engage in costly imitation. If investors can commit to a re-hiring rule with profitable deviations, they optimally choose to re-hire any manager with any positive return. The model results match empirically observed flows in mutual funds.

In the third chapter I investigate limited liability, unlimited liability, and bankruptcy for entrepreneurs with credit constraints. Limited liability offers a strictly larger choice of contracts than unlimited liability for fully liquid and illiquid entrepreneur wealth. Bankruptcy is redundant with limited liability under liquid wealth, although it is useful for entrepreneurs when access to limited liability is exogenously restricted or costly. Making bankruptcy more generous with illiquid wealth strictly reduces the choice of viable contracts. The model results match empirical findings that lower barriers to entry boosts entrepreneurship while generous bankruptcy has mixed impacts on entrepreneurship.

Contents

List of Figures

List of Tables

1 Betting with Beta

1.1 Introduction

While mutual fund literature has long debated whether *skilled* managers who consistently outperform the market exist, it has been thoroughly documented that *unskilled* managers exist and have persistently poor performance. These managers can be found in the left tail of benchmark adjusted return or alpha distributions sorted over long horizons, and they persistently destroy value for their investors. Left tail funds suffer penalties via outflows but not as much as their right tail counterparts gain from inflows, and some left tail funds survive for lengthy periods of time while underperforming their benchmarks and generating negative alpha (Cuthbertson et al. (2010)).

The central hypothesis of this paper is that (i) bad managers gamble (increase the volatility of their portfolio returns), and (ii) these managers can be detected in the left and right tails of the *short*-term return distribution. The study does not take a stance on why poor quality managers increase their volatility - for example, it could be a rational response to gamble with their own private information of their lack of skill combined with the observable convex flow-performance relationship, or it could be behavioral overconfidence. Regardless of the motivation for gambling, if bad managers have a higher volatility of their payoffs, it follows that bad managers can be identified by looking at both the right and left tails of the short-term return distribution. Bad managers do not need months or years to reveal their type by consistently underperforming their benchmarks; by doing suspiciously well or suspiciously poorly over short horizons, they can be identified. Having a large negative return is naturally a bad signal of manager quality, but counterintuitively, above some point in the right tail of the return distribution, you will also encounter more bad managers than good.

To demonstrate, suppose that bad managers underperform their benchmark, having on average negative excess return (R^{ex}), while good managers on average outperform their benchmark. The 'good' managers could also have zero average return - for this hypothesis, it doesn't matter if *skilled* managers exist or not, so long as *unskilled* managers exist and underperform the average fund. Both managers draw returns from a student T distribution, with the bad manager having a lower mean and a higher standard deviation. A graphical demonstration of the intuition that bad managers can reveal their type by doing too well is shown in Figure 1.

Figure 1 demonstrates how bad managers can reveal themselves by achieving returns in either tail. Intuitively, earning extreme negative returns reveals a manager to be poor quality, as it is both an indicator of having a low mean and a high standard deviation. But above a certain point in the right tail, return is *also* a bad signal; above R^*, there are more bad managers than good.

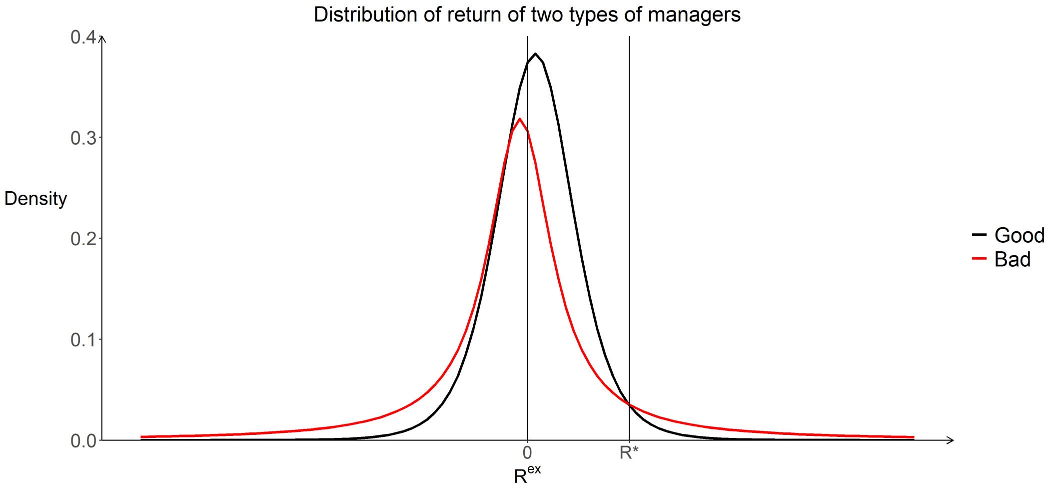

Figure 1. Bad and good quality managers both draw monthly portfolio returns from a student T distribution. The bad manager has a lower mean but a higher standard deviation of returns.

To investigate the hypothesis, I sort mutual funds into buckets based on their return in excess of the market. The buckets are formed based on absolute instead of relative performance, which allows for the possibility that during volatile times gambling can have a higher payoff than during normal times. Once buckets are formed I measure the future performance of mutual funds in each bucket over one to five year horizons. I examine the funds' cumulative return in excess of the market, their Sharpe Ratio, and most importantly to control for style, their alpha.

Measuring the performance of right and left tail funds brings a unique set of challenges. Funds with extreme returns have extreme flows, and those flows can change how they balance their portfolio: Lou (2012) finds that funds spend 62 cents of every dollar of inflows on expanding existing positions, leaving the remainder for new positions. Especially for funds that are suspected of 'gambling' - intentionally choosing a high volatility of payoffs - their beta exposures can change rapidly. The typical way of measuring funds' beta exposures, regressing fund returns on factors over one to five year horizons, is not suitable for funds with extreme returns and flows.

My novel solution to the problem is to construct fund betas from their portfolio holdings. Betas for stocks are found by regressing stock returns on factors, and the resulting stock betas are used to form portfolio betas of mutual funds. Since funds report holdings quarterly for most of the sample, I am able to track funds' betas with higher accuracy than the traditional method, especially over short horizons. The methodology allows insight into how often funds change their beta, an insight which has implications beyond the initial question of this paper. I document that funds, especially in the tails of the return distribution, change factor exposures much more frequently than is suggested by the traditional regression of fund returns on factors.

The downside of this approach for measuring performance is that it introduces a layer of selection bias into results. Funds that are larger, have better returns, and have a longer lifespan are more likely to have holdings data that are linked to fund returns. Additionally, measuring mutual fund performance over future horizons mechanically introduces survivorship bias; I can only measure the one-year average monthly alpha of funds that survive for at least one year after they are sorted into their buckets. These two biases mean that results should be interpreted as upper bounds of the true performance of funds.

Despite the biases discussed above, the results in the right tail of mutual fund return conform to the hypothesis; above a threshold, contemporaneous excess return is a negative predictor of future return, Sharpe ratio, and alpha over one to five year horizons. The initial high, positive inflows for these funds reverse and become on average negative after one year. The threshold for R^* is high - the underperforming right tail is formed of funds that have beaten the market by at least 15% in the prior month. Above this threshold funds have statistically and economically significant negative future performance from one to five year horizons.

Measuring funds in the extreme left tail delivers the opposite results - the extreme left tail is predictive of *positive* performance across the same horizons. Since results are an upper bound on performance due to the matching between return and holdings and most results are not statistically significant, it's unclear if these funds are actually performing better than their peers. I discuss results from other papers which document that left tail funds often change strategy and fire their previous manager as a possible explanation for genuine improvement in the funds' performance. Ironically, right tail funds underperform their left tail counterparts - interpreting this in the context of the original hypothesis, gamblers who get *unlucky* are revealed as bad quality and fired, while gamblers who get *lucky* are able to continue their value destroying strategies for some time.

After examining funds' performance, I investigate if the fund shows evidence of 'gambling'. Right tail funds don't show any evidence of excessive leverage; the typical US mutual fund has positive cash positions. Maximum leverage is regulated through the SEC, but most funds are well below the legal ceiling, and use cash holdings to facilitate liquidity transformation (Chernenko and Sunderam (2016)). Right tail funds do have higher turnover ratios and portfolio concentration than the average mutual fund, although only the result on turnover ratio is statistically significant. The largest difference in the behavior of right tail funds comes from their beta exposures. Right tail funds have not only high average factor exposures - which could be due to style - but also have much higher standard deviation of factor exposures. Funds adjust their factor exposures much more rapidly than other funds in the hope of getting a lucky return realization. In addition to having a higher standard deviation of betas over their lifetime compared to the average fund, in the holdings observation just before their return is realized, right tail funds have higher changes in factor exposures even compared to their own long-run average.

Identifying a channel for gambling through beta allows me to potentially identify other gamblers who weren't (un)lucky enough to receive an extreme return. I estimate how many more funds are gamblers than are revealed solely by receiving an extreme return. I find that funds which have strategies that are at least as extreme and transient as the average right tail fund also have statistically and economically significant negative performance in the year following their identification as gamblers.

One important caveat for the results is that they are for a date range that does not include the dot com bubble. Because funds are sorted by absolute instead of relative return, the buckets of extreme return ranges are full of observations during the dot com bubble: for example, about 90% of fund-date observations that beat the market by at least 15% within the past month during 1990-2023 come during the dot com bubble. By including the dot com bubble in the sample results are driven solely by fund performance during those years. Sorting by absolute return over the dot com bubble selects almost entirely tech funds into the right tail when the bubble is rising and into the left tail when it has popped. The purpose of the original hypothesis is to characterize fund manager behavior during normal times, not to study fund performance during bubble episodes. In the robustness section, I examine how results differ when the dot com bubble in included.

This paper is related to several strands of literature. The question about whether skilled mutual fund managers exist has been debated for some time. Seminal papers include Carhart (1997), which finds that when the momentum factor is controlled for in performance mutual fund managers have no significant alphas, and Berk and Green (2004), which finds that managers have skill, but that positive alphas are eliminated through positive flows through decreasing returns to scale. Cuthbertson et al. (2010) is a literature review that summarizes many papers in this vein and concludes that evidence of mutual fund *underperformance* in the left tail of alpha is well-established, but that evidence of truly skilled managers is sparse to non-existent. In this paper I contribute a novel method to identify unskilled managers; they can be found in the left tail of sorting funds by alpha but also in the right tail of the short-term return distribution. My result suggests that managers may be identified before they have destroyed value for several years in a row, which is important for investors.

Papers that find some (albeit few) managers have skill conclude that skilled managers should be found by looking in the extreme right tail of the alpha distribution. For example, Barras et al. (2010) suggests a very similar conceptual model to what I display in Figure 1, albeit with good and bad managers[1] possessing the same vari-

[1] Barras et al. (2010) finds that most managers are zero-alpha and discuss three distributions: bad managers, zero-alpha managers, and good managers. To relate it to my findings, I will discuss only the distributions of good and bad managers.

ance of their alpha realizations. When good managers have a higher mean of alpha and the two types have the same variance, the further you look in the right tail of long-run performance, the more *good* managers you find; the less ‘false discoveries’ are made of bad managers that got lucky. In contrast, my hypothesis is that bad managers have a lower mean but higher variance of return. I find that extremely high short-term return is indicative of poor future performance.

Previous authors have decomposed mutual fund manager skill into distinct categories such as stock picking versus market timing. The results of this paper indicate that bad mutual fund managers attempt to time their factor exposures but are unsuccessful. Market timing has been studied in theory (Admati and Ross (1985)) as well as empirically in a seminal paper by Daniel, Grinblatt, Titman, and Wermers (Daniel et al. (1997)). Similar to the results of this paper, the authors of Daniel et al. (1997) find that funds exhibit no market timing abilities.

This paper also studies the behavior of investors and the flow-performance relationship. In the ‘Flow’ subsection of results I recreate the classic convex flow-performance relationship first identified in Chevalier and Ellison (1997). I find that investors are fooled by the extreme right tail short-term performance of unskilled managers, but eventually flows reverse as the managers begin to underperform across multiple metrics. Investors’ lack of sophistication in initially flowing into these gambling funds is discussed more recently in Ben-David et al. (2022), which finds that investors mechanically flow into high performing funds instead of investigating fund performance as rational, sophisticated investors seeking alpha.

One contribution of this paper is to investigate how mutual fund managers gamble. A seminal paper studying fund manager gambling is Brown et al. (1996), which finds that mid-year funds in the bottom half of the return distribution become relatively more risky while mid-year funds in the top half of the return distribution become relatively more safe. The idea that bad mangers (which are more likely to belong to the bottom half of the return distribution over any period) gamble is central to this paper. In contrast to their results on safety, I find that once managers get lucky they continue to gamble, eventually underperforming the market and their beta exposures by large margins for years after their initial spectacularly high return. Wermers (2012) find that funds with more consistent style have higher performance, which matches my findings that poor quality managers gamble by rapidly changing their beta exposures.

The novel methodology of this paper in measuring factor exposures using the holdings of funds is most similar to Kacperczyk et al. (2008). In the original paper authors compare the return of mutual funds against the passive return of their disclosed holdings, finding that funds with a positive return gap have high future performance while funds with a negative return gap are unskilled. In measuring factor exposures from the holdings of funds I take the methodology one step further by controlling for factor exposures which impact fund return. I use the results on fund factor exposures to investigate if funds in the right tail of returns are skilled of unskilled.

The remainder of the paper is organized as follows. Section 2 discusses data methodology. Section 3 presents results on the future performance of funds sorted by volatility and by return. Section 4 investigates how right tail funds are gambling. Section 5 discusses robustness and Section 6 concludes.

1.2 Data

Data on returns and fund characteristics are from CRSP, while holdings of long equity positions comes from Thomson Reuters. Returns from CRSP are used to measure return predictability, while holdings data from Reuters allows insight into the funds' factor exposures.

Starting from the entire CRSP universe of funds I first drop ETFs and index funds. Using CRSP's style code I further filter the sample to active, domestic (US), non-sector equity mutual funds. From 2003-2023, about 2.3 million share class - month level observations fit this classification.

Using WFICN from MFLink, I map share class level observations into fund level observations. If multiple share classes exist within a fund, the fund's total net assets (TNA) is the sum of share class TNA, and the fund's age is the maximum age of any share class. All other fund-level variables - for example, return - are taken as the value weighted average of share class level information. From 2.3 million share class-month observations I obtain 698,339 fund-month observations.

After dropping observations based on the fund's classification, I apply filters to eliminate extreme outliers from the analysis. First, fund-month observations with TNA $<$ 5 million are dropped to ensure the results are not driven by small funds. Second, monthly returns greater than 100% or less than -50% are dropped. Although this paper studies returns in the tails, the data present a small number of extreme outlier returns (less than 50 observations) that are too large to be believable.

Similarly, since flows are also studied, funds with monthly flows above 100% or less than -50% are dropped. 659,734 fund-month observations remain after both of the prior filters are in place (38,605 dropped from the full sample). These observations form the core sample which I will use to study predictability of performance.

In addition to return and size information, expense and turnover ratio are downloaded from CRSP. The expense ratio allows for the calculation of the fee the fund charged - the object of this paper is be pre-fee return, so fees need to be added back to the raw return from CRSP[2]. Both expense and turnover ratios will be used as controls in regressions to ensure results are not driven by funds with high fees or

[2] For funds missing expense ratio information, no fee is added back into returns. As the fee is typically very small compared to the realization of the market this does not distort results.

funds that trade (in)frequently.

Holdings characteristics from Thomson Reuters are matched to CRSP using WFICN. Holdings are reported every quarter for the typical fund in the sample[3], while returns are reported on a monthly basis. I merge holdings with returns by assuming holdings are 'forward looking': holdings match with future returns instead of past returns. For example, a hypothetical end of March holdings observation starts to match with returns on April and continues to match until holdings are reported again. Fund holdings have been found to more reliably indicate future holdings than past holdings - see, e.g. Kacperczyk et al. (2008). Fund holdings are matched to returns at a maximum horizon of one year, which can occur when one to three holdings observations are missing from Reuters.

Several filters are applied to ensure the match of returns from CRSP and holdings from Reuters is high quality. First, holdings observations with fewer than 20 stocks are dropped. With few exceptions, mutual funds are not allowed to hold more than 5% of their assets in any individual stock, so any observed holdings with fewer than 20 tickers is incomplete. Second, I check that the sum of holding positions is close to the fund's TNA. A fund's stock holdings must be at least 80% of the fund's recorded TNA on the same month, and must be below 200% of the fund's TNA. The generous upper limit is to account for the fact that mutual funds may be missing one of the fund's share classes on CRSP, in which case the fund's TNA will be lower than their holdings. However, missing a share class in holdings data does not rule out the fund as a valid candidate to study return predictability or to study the beta exposure of their portfolio.[4] In each case, if a fund does not have a high quality match with Reuters data, I retain the observation to study return predictability but set all holdings variables to missing, as if there was no match with their WFICN in the holdings data.

364,266 of the original 659,734 fund-month return observations have high quality matches to holdings data. This is after 54,093 observations were set to missing for having fewer than 20 stocks, 76,003 observations were set to missing for having stock values too low relative to their TNA, and 15,293 observations were set to missing for having stock values too high relative to their TNA.

For the funds with a high quality holdings match I record the number of shares they hold as well as a measure of portfolio concentration, the Herfindahl-Hirschman Index or HHI (Rhoades (1993)). HHI is found by calculating

$$HHI_{i,t} = \sum_{j=1}^{J} w_j^2 \tag{1}$$

[3] Since 2003 funds are required to disclose quarterly holdings to the SEC, although some voluntarily report holdings more frequently.

[4] Results are robust to varying the 200% upper threshold for matching TNA reported in CRSP to the sum of holdings reported in Reuters.

For all stocks j in the fund's portfolio. HHI increases in portfolio concentration.

In order to measure a fund's alpha, I obtain beta exposure of individual stocks from the monthly stock returns on CRSP. Stock exposures for three and seven factors (respectively) are estimated using a 60 month rolling window regressions of stock returns on factors, where the factors are obtained from Kenneth French's website (French (2025)). R_{jt} denotes stock j's excess return in month t:

$$R_{j(t,t-60)} = \beta_{jt}^{mkt} MKT_{(t,t-60)} + \beta_{jt}^{SMB} SMB_{(t,t-60)} + \beta_{jt}^{HML} HML_{(t,t-60)} \tag{2}$$

$$\begin{aligned} R_{j(t,t-60)} = \beta_{jt}^{mkt} MKT_{(t,t-60)} + \beta_{jt}^{SMB} SMB_{(t,t-60)} + \beta_{jt}^{HML} HML_{(t,t-60)} + \\ \beta_{jt}^{UMD} UMD_{(t,t-60)} + \beta_{jt}^{RMW} RMW_{(t,t-60)} + \beta_{jt}^{CMA} CMA_{(t,t-60)} + \beta_{jt}^{STR} STR_{(t,t-60)} \end{aligned} \tag{3}$$

Results for alphas will be reported on both three and seven factor specifications. In addition to the standard market, size, and value factors, the momentum (UMD), profitability (RMW), and investment (CMA) factors are included (Fama and French (2015), Carhart (1997)). The final factor, Short Term Reversal (STR), is included to ensure alphas are not driven by mechanical stock price reversion. Since the paper studies funds that by definition have received extreme returns, controlling for mechanical short-term reversion is important for robustness of results.

Stock betas are matched by ticker and date to mutual fund holdings from Reuters. Fund i's betas at time t are mechanically found from each of its stock holdings (j):

$$\beta_{it} = \sum_{j=1}^{J} w_j \beta_j \tag{4}$$

Factor exposures are adjusted by multiplying them by (1− cash) when data from CRSP on funds' cash holdings are available. This is to reflect that cash holdings have a beta of zero. If funds do not have data on cash holdings the betas do not receive an adjustment. Alpha results are robust if this adjustment is removed for all funds. Portfolio betas are found using this formula for both three and seven factor exposures. In the rare event that a stock has insufficient returns data to compute betas - for example, if the stock just had its IPO - the stock is excluded from the portfolio betas, and the weights of other stocks are readjusted accordingly.

The tables below include summary statistics for data taken from CRSP and Reuters.

Table 1
Summary statistics - CRSP

	Mean	Median	Standard Deviation	N
Total observations	659,734	-	-	-
Number of distinct funds	5,818	-	-	-
Number of distinct months	251	-	-	-
Number of months per fund	113.40	97	81.37	-
Pre-fee Return	0.0083	0.0116	0.0480	659,734
TNA (millions)	1577.73	262.8	6332.61	659,734
Age (years)	13.58	10	12.95	659,500
Expense Ratio (%)	1.06	1.05	0.48	511,187
Turnover Ratio (%)	0.74	0.50	0.98	510,682
Percent cash (%)	2.70	1.31	4.59	587,520

Table 1 displays summary statistics of variables obtained from the CRSP database. Total observations displays the number of fund-month observations after data filtering. The 'Age' variable is taken as the maximum age of any share class and the 'TNA' variable is the sum of TNA of each share class, while other variables are found as the weighted average of share class level variables, with share class TNA as weights.

Table 1 summarizes variables from the CRSP database. As previously stated, the 'Age' variable is the maximum age of any of the fund's share classes, while the Total Net Assets ('TNA') variable is the sum of share class TNA. All other fund variables are weighted averages of share class variables, where weights are formed by the TNA of each share class.

Table 2 summarizes variables from the Thomson Reuters database. The first row, holdings observations, reports the number of holdings reports that have high quality matches with returns data. The second row, fund-date observations, records the total number of months where returns observations have high quality holdings data. Holdings observations are typically matched with multiple months of returns data. Most funds report regular quarterly holdings: only 10,760 of 114,275 total fund holdings observations are more than a quarter from the previous holdings observation, while 4,111 holdings observations report holdings after one or two months. These funds are voluntarily reporting holdings at a higher frequency than is required by the SEC. As stated previously, holdings are matched with returns on a forward looking basis: for example, end of March holdings observations are matched with returns from April through June if the fund reports quarterly holdings, with end of June holdings beginning to match with July returns.

To compare different methodologies of calculating betas, Table 3 reports summary statistics of betas calculated in the conventional fashion as well as my novel method of calculating betas from holdings. The conventional method finds fund factor exposures by calculating

Table 2
Summary statistics - Reuters

	Mean	Median	Standard Deviation	N (holdings)
Holdings observations	114,275	-	-	-
Fund-date observations	364,266	-	-	-
Months between holdings	3.31	3	2.67	114,275
Number of shares	138.03	81	233.60	114,275
HHI	0.038	0.0232	0.441	114,275
3F MKT	1.054	1.043	0.159	113,201
3F SMB	0.257	0.156	0.377	113,201
3F HML	0.047	0.055	0.030	113,201
7F MKT	1.018	1.014	0.131	113,201
7F SMB	0.264	0.159	0.376	113,201
7F HML	0.048	0.049	0.243	113,201
7F UMD	-0.051	-0.052	0.110	113,201
7F RMW	0.015	0.038	0.240	113,201
7F CMA	-0.095	-0.070	0.264	113,201
7F STR	0.0002	0.0002	0.001	113,201

Table 2 displays summary statistics of variables obtained from the Thomson Reuters database. Holdings observations displays the number of unique holdings observations with high quality matches to returns, while fund-date observations displays the number of monthly returns observations matched with holdings data. Each holdings observation is matched with multiple months of returns, as is explained in text below. The number of shares and HHI are calculated directly from holdings data, while the factor exposures are found by regressing each stock on factors with a 60 month rolling window to obtain stock factor exposures. Portfolio betas are found by taking the weighted average of stock betas within each fund - holdings observation. Note that the number of holdings observations with beta exposures is slightly smaller than the total number of holdings observations: this occurs rarely when a fund has fewer than 20 stocks with 60 months of returns data. (Holdings observations with fewer than 20 *total* stocks in holdings have already been dropped.) As the next section explains, stocks without 60 months of consecutive returns observations are not assigned factor exposures.

$$R_{i(t,t-w)} = \beta_{it}^{mkt} MKT_{(t,t-w)} + \beta_{it}^{SMB} SMB_{(t,t-w)} + \beta_{it}^{HML} HML_{(t,t-w)} + \beta_{it}^{UMD} UMD_{(t,t-w)} + \beta_{it}^{RMW} RMW_{(t,t-w)} + \beta_{it}^{CMA} CMA_{(t,t-w)} + \beta_{it}^{STR} STR_{(t,t-w)} \quad (5)$$

For all funds i at each month t with rolling time window w. In order to investigate how quickly funds are able to change their betas I first calculate how much betas have changed from the last three months within fund for 12 and 60 month rolling time windows, in addition to my measure which creates fund betas from holdings. This is found by taking the absolute value of the change of funds' exposures for each methodology: for example, for SMB, I calculate $|SMB_{i,t} - SMB_{i,t-3}|$. I then calculate the within-fund average change to get one observation per fund. Finally, Table 3 reports the average across funds. Column 1 shows the results for betas calculated by regressing fund returns on factors for a 12 month window, column 2 reports results from calculating factor exposures from holdings data, and column 3 shows results for betas calculated by fund returns on factors for a 60 month window.

Table 3

Mean absolute changes in betas

	12 month window	From holdings	60 month window
3F MKT	0.090	0.049	0.015
3F SMB	0.127	0.058	0.021
3F HML	0.138	0.071	0.024
7F MKT	0.158	0.048	0.017
7F SMB	0.258	0.059	0.023
7F HML	0.286	0.072	0.028
7F UMD	0.219	0.040	0.017
7F RMW	0.406	0.089	0.035
7F CMA	0.412	0.098	0.040
7F STR	0.0021	0.0004	0.0002
N	457,362	482,073	286,326

Table 3 compares the mean absolute change in betas over three months between the three methods of beta formation. In the first column, I show the mean absolute change in betas over three months with the conventional method of calculating betas by regressing stock returns on factors over a 12 month window. The second column shows the mean absolute change in betas over three months when my method is utilized, computing portfolio betas from holdings data. In the final column, I show the mean absolute change in betas over three months with the conventional method of calculating betas by regressing stock returns on factors over a 60 month window.

In Table 3, the 12 month rolling window betas are the most volatile, while the 60 month rolling window betas are least volatile. The beta measures from holdings lie in the middle. Note that although matching returns data to holdings drops some observations, it is actually the *largest* sample of factor exposures between the three methods. One benefit of calculating factor exposures from holdings is it does not require 12 or 60 months of prior consecutive returns in order to form factor exposures, which reduces survivorship bias and increases the sample size of fund alphas.

In addition to the conventional method of regressing fund returns on factors having fewer observations than finding fund exposures from holdings, I argue that the 12 month rolling window betas overreact to changes in strategy for right tail funds, while the 60 month rolling window betas underreact to changes in strategy. Even if an intermediate time window was used - for example, 36 months - I could potentially match the mean absolute changes in betas calculated from holdings but the window would be too long to react quickly to changes in strategy. In other words, for time windows short enough to pick up changes in strategy, betas appear to overreact to these changes in strategy. In Section 4 I show that right tail funds vary betas considerably between holdings observations, motivating my search for a method of beta formation which adjusts quickly to changes in strategy.

Figures 2 and 3 graphically demonstrate my argument for using betas from holdings to measure the performance of funds. Each displays a single factor exposure of a single fund over time. The example funds used are in the right tail of the short-term performance distribution for at least one month of their lifespan, and are the target for my analysis in this study. In the first figure, one benefit of estimating betas from holdings becomes immediately clear: beta estimation from holdings becomes available to measure before the other two methodologies. The fund does not need to exist for 12 or 60 months in order to get the first beta exposure, reducing survivorship bias in my analysis. The 12 month rolling window appears extremely noisy; it is almost always more extreme than the other two estimates, often by large magnitudes, and changes signs rapidly over short horizons. Note that the example fund has an extreme short-term return realization at some point in its lifetime, meaning it has a high volatility of returns. The 60 month rolling window is much less volatile but misses changes in strategy. In particular for this fund, around 2006 they quickly increased its exposure to SMB, which was only reflected around 2009 in the 60 month rolling window beta. Also, in the period from 2008 to about 2015, the fund was rapidly changing its SMB exposure, while the 60 month rolling window only displays the average strategy. Overall its SMB exposure declined, which was eventually reflected in the 60 month rolling window, but there were also several periods where the fund bought stocks with very high exposure to SMB which were quickly re-sold.

In the second figure, HML exposure of a second fund is tracked with a similar pattern. HML exposure is able to be formed more quickly by estimating factor exposures from holdings than with either rolling window estimation, allowing insight to the fund's beta before it reaches an age of one or five years. The beta from holdings decreased rapidly and changed quickly between the 2003 and 2008 periods, which the 60 month rolling window estimation takes several years to reflect: the estimated exposure can be observed slowly and continuously falling until it matches the fund's strategy revealed by holdings around 2008. As in Figure 2, Figure 3 shows that the 12 month rolling window beta sometimes follows the general pattern of the other methodologies but is extremely noisy and changes signs rapidly.

Another way to interpret the novel way that I calculate exposures is to consider it an extension of Kacperczyk et al. (2008). The authors of Kacperczyk et al. (2008) measure fund performance by measuring the returns of a passive portfolio of mutual funds' holdings against the their realized returns. In my methodology I'm taking the idea one step further: rather than measuring if mutual funds have a higher raw return than their passive portfolio I'm measuring if funds outperform their reported passive factor exposures in holdings.

1.2.1 Variable definitions

The return variable of interest is the pre-fee market adjusted return at each fund-month, calculated as follows:

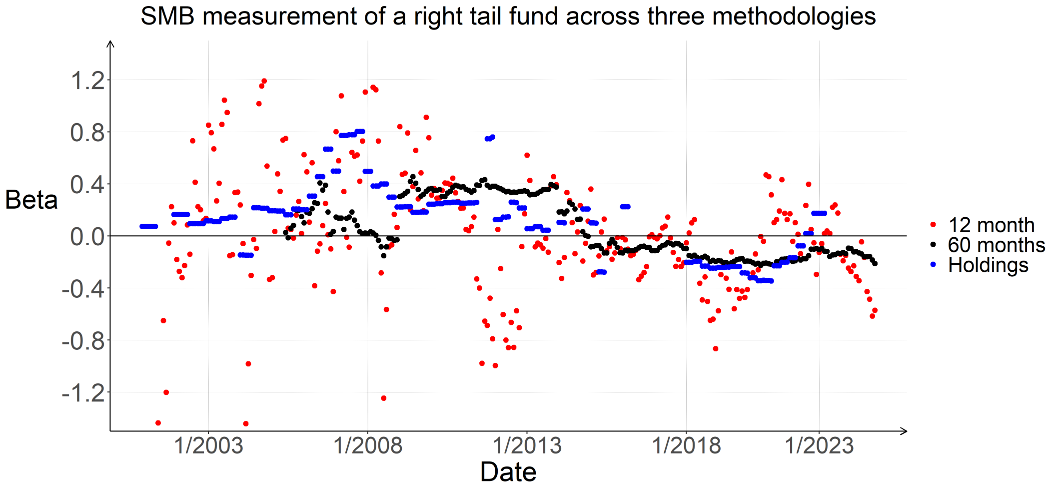

Figure 2. This figure displays the SMB factor beta of an example right tail fund across three different methodologies. In red, I use the regression of fund returns on factors with a 12 month rolling window. In black, I use the regression of fund returns on factors with a 60 month rolling window. In blue, I measure fund betas from their holdings.

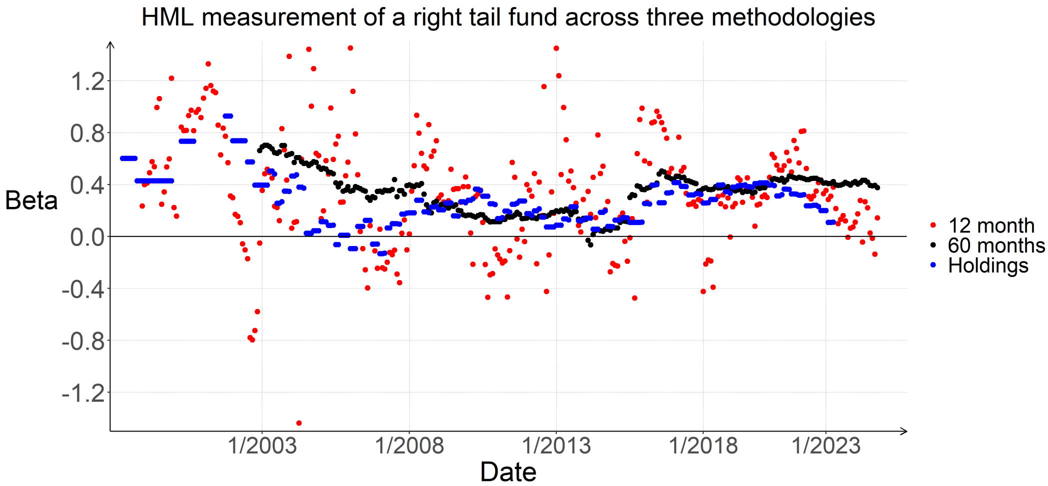

Figure 3. This figure displays the HML factor beta of an example right tail fund across three different methodologies. In red, I use the regression of fund returns on factors with a 12 month rolling window. In black, I use the regression of fund returns on factors with a 60 month rolling window. In blue, I measure fund betas from their holdings.

$$R^{ex}_{i,t} = R_{i,t} + fee_{i,t} - R^{S\&P}_t \tag{6}$$

For simplicity, for the remainder of the paper, I will refer to $R^{ex}_{i,t}$ as 'excess return'. $R_{i,t}$ denotes the return of the fund in a month as reported by CRSP, which is reported net of fees. The fee is added back to the return using the expense ratio. This study focuses on pre-fee returns because the hypothesis is about manager skill, not about if earnings of funds accrue to investors. I subtract the return of the S&P 500 to adjust for market returns in the current month. This is identical to assuming that all funds use the S&P 500 as their benchmark. The first part of the study focuses on predictability of excess return. To further evaluate skill, risk adjusted returns will be examined via the fund's alpha, which controls for fund style.

There are two ways of measuring mutual fund flow in the current literature. The first is a standard percent change calculation:

$$F_{i,t} = \frac{TNA_{i,t} - (TNA_{i,t-1}(1 + R_{i,t}))}{TNA_{i,t-1}} \tag{7}$$

The second adds a $(1 + R_{i,t})$ term to the denominator:

$$F_{i,t} = \frac{TNA_{i,t} - (TNA_{i,t-1}(1 + R_{i,t}))}{TNA_{i,t-1}(1 + R_{i,t})} \tag{8}$$

This second equation puts units in time-t dollar amount to be consistent with the numerator. It guarantees that a fund which loses all of its capital has a flow of -1, while the previous equation can potentially have flows of less than -1.

Neither is satisfying for the purposes of this study. The first method of measuring flows (7) is consistent with flows that enter at the end of the following month after returns are realized. The realized returns in the subsequent month do not grow or shrink flows that enter during the month. The second method (8) is consistent with flows that enter at the beginning of the month, and returns fully impact these additional cashflows. The methods are roughly equivalent when return realizations are small, but diverge when return realizations are large in magnitude. The former tends to overstate flows when the fund is doing well and understate flows when the fund is doing poorly, while the latter understates flows when the fund is doing well and overstates them when the fund is doing poorly. The tail funds of interest for this study have by definition received large returns (either positive or negative), are likely to have large flow responses to the return realizations, and may have large return realizations in following months.

When returns are large, the difference between these methods can be sizable. Suppose a fund reports a TNA 10% above the last month's TNA but reports a -20% return. The first method would compute flows of +30%, while the second method would compute flows of +37.5%.

For my purposes I adopt a compromise, assuming half of flows are fully impacted by returns:

$$F_{i,t} = \frac{TNA_{i,t} - (TNA_{i,t-1}(1 + R_{i,t}))}{TNA_{i,t-1}(1 + 0.5R_{i,t})} \tag{9}$$

The $(1 + R_{i,t})$ term still appears in the denominator, but it has been shrunk to reflect that the flows do not enter immediately at the beginning of the subsequent month. In the absence of a study documenting when mutual fund flows occur within a month, this compromise is more consistent with flows and returns that accrue uniformly through the month than either of the methods used in prior literature. However, results on flows are robust to using any of the three methods explained above.

1.2.2 Selection bias of variable inclusions

This study uses several subsets of data when examining the future performance of mutual funds. First, CRSP returns by themselves have been shown to be upward biased (Elton et al. (2001)). Second, by studying future performance of funds, I mechanically introduce look-ahead bias that is more severe with longer horizons. For example, a fund must continue to exist for a year following bucket formation in order to have a 12 month cumulative excess return. Additionally, two sets of controls are considered for all regression specifications. In all specifications, transformations of the fund's size and age are included as controls. In every month, the median log(age) of each fund is found. The age control variable is then each individual fund's log(age) minus the median log(age) in the month. This transforms the age control into a stationary variable; without the transformation, the average log(age) is increasing throughout the sample period, which is problematic for regression analysis. An identical transformation with the fund's size ensures that size is a stationary variable through the sample period.

A second specification adds expense and turnover ratio in control variables to ensure results are not driven by funds that have high fees or funds that trade often. Since some observations have return, size, and age information but not an expense or turnover ratio, these controls do slightly decrease the sample size. The two sets of controls are therefore chosen so that the first utilizes as many observations as possible, while the second controls for common characteristics. Requiring funds to have expense and turnover ratio controls positively selects funds; these funds are larger and have higher excess return than other funds in the sample. The upward bias on size and return is especially strong when considering funds in either extreme tail of the return distribution.

Table 4 documents the selection that occurs by regressing fund size and fund performance on a dummy variable for a fund having non-missing expense and turnover ratio data. That is, I run

$$y_{i,t} = \alpha + \beta_1 \mathbb{1}^{ER,TR} + timeFE_t + \epsilon_{i,t} \tag{10}$$

With $y_{i,t}$ taking values of (i) fund TNA and (ii) excess return over the following year. In order to demonstrate the full-sample bias gets stronger when assessing the performance of tail funds, I run the regression in Equation 13 multiple times; for each left hand side variable I run one regression with the full sample, a second regression with extreme left tail funds (with contemporaneous excess return of less than -15%), and a third regression with extreme right tail funds (with a contemporaneous excess return of greater than 15%). The rows of the table specify the left hand side variable, and the columns specify the sample. The estimates of β_1 are reported in the table below.

Table 4 shows that the addition of the expense and turnover ratios as control

Table 4
Selection in controls

	Full sample	$R^{ex} \leq -15\%$	$R_{ex} > 15\%$
TNA	900.287***	1121.691***	193.256
$R^{ex}_{t\in 1,12}$	1.033***	6.810**	11.618**
Time FE	yes	yes	yes

Table 4 compares the total net assets and cumulative twelve month excess return between subsets of observations. In the first column, the average of TNA and return are computed for funds which have expense and turnover ratios. This should be compared with the full sample averages reported in Table 1. In the second column, the average of these variables for extreme left tail performers with non-missing expense and turnover ratio is shown. In the third column, the average of these variables for extreme right tail performers with non-missing expense and turnover ratio is shown.

variables add a layer of selection bias to the results. Funds that have both expense and turnover ratio are on average 900 million larger and earn 1% cumulatively more than their peers over the year following bucket formation for the full sample. The bias towards larger funds is much higher when just the extreme left tail is considered but not statistically significant in the extreme right tail. For the extreme left tail, the average fund with non-missing expense and turnover control variables has a 6.810% higher cumulative excess return over the following year. The bias increases to 11.618% when the regression is run in the sample of extreme right tail funds.

Results in the tails that are still statistically significant with the addition of expense and turnover ratio controls are very robust, but if adding these control variables diminishes the statistical significance, it should not be interpreted as overturning the result. Regressions with these control variables present a positively biased upper bound of performance. In the full sample, 158,888 fund-month observations are lost when expense and turnover ratios are included as control variables.

In addition to requiring expense and turnover ratios in some specifications, any regression with monthly alpha as a left hand side variable requires funds to have information for factor exposures, which requires them to have a high quality match with Thomson Reuters holdings data. Table 5 runs an identical regression to Table 4 replacing the dummy variable to equal one if the fund has a high quality match with Reuters and zero otherwise.

The selection bias introduced by requiring a high quality match with holdings is significant in the full sample. Funds that match with holdings data are on average 216 million larger in size and perform about 1.8% better in cumulative excess return over the following year than other funds. When just considering the extreme left tail the bias is not high or significant: funds with holdings information are about 101m larger than other extreme left tail funds without holdings information, and unusually, funds with holdings information perform worse than their peers without holdings information, albeit without statistical significance. For the extreme right tail the

Table 5
Selection in holdings

	Full sample	$R^{ex} \leq -15\%$	$R_{ex} > 15\%$
TNA	216.662***	101.639	340.516*
$R^{ex}_{t\in 1,12}$	1.792***	-0.647	14.270***
Time FE	yes	yes	yes

Table 5 compares the total net assets and cumulative twelve month excess return between subsets of observations. In the first column, the average of TNA and return are computed for funds which match with holdings data from TR. This should be compared with the full sample averages reported in Table 1. In the second column, the average of these variables for extreme left tail performers with non-missing holdings data is shown. In the third column, the average of these variables for extreme right tail performers with non-missing holdings data is shown.

inclusion of holdings information introduces a strong upward bias. Funds with holdings information are about 340m larger, although the result is only significant at the 10% threshold. However, these funds have an average of 14.270% higher cumulative excess return in the following year than extreme right tail funds without holdings information.

One final control variable considered for this study is the funds' Morningstar rating. However, funds with Morningstar ratings are extremely positively selected, especially when ratings are required alongside holdings information and expense and turnover ratios. The inclusion of Morningstar controls is especially problematic for funds in the tails. When Morningstar ratings are added as as a right hand side variable nearly all estimates for performance - return, alpha, etc. - in any bucket of the return sort become positive relative to the omitted group. A summary of the selection bias introduced by including Morningstar ratings as a control is provided in the Robustness section.

As explained by the introduction, results - especially alpha results requiring high quality holdings information that use the full set of controls with turnover and expense ratios - should be considered an upper bound of fund performance.

1.3 Results

The results section will begin by presenting regressions studying the predictability of excess return, Sharpe Ratio, and alpha of funds sorted into buckets based on volatility and return. The central hypothesis of this paper is that bad managers gamble, i.e. have a higher volatility of portfolio return. I investigate if these bad managers reveal their type by having (1) high volatility over the past year or (2) by having extreme positive or negative returns.

1.3.1 Sorting funds by volatility

The original hypothesis predicts that funds with high volatility or return will have negative future performance across a number of performance measures. The simplest implementation of this hypothesis is sorting funds into buckets based on their volatility in the year before their return observation and looking at measures of future performance: cumulative excess return, Sharpe Ratio, and three and seven factor alphas. In this section I focus on performance in the year following the funds' return realizations (t+1 to t+12). The robustness section discusses performance over other horizons.

One novel contribution of this paper is sorting funds into buckets based on absolute instead of relative thresholds. Although sorting by relative performance has the merit of filtering out common shocks between funds, sorting by absolute performance allows for the possibility that volatility about a threshold - or outperforming the market at some extreme threshold - could be a signal of future performance. In other words, funds do not populate the most extreme buckets of volatility and performance in every month, but perhaps the extreme levels of volatility or performance that are only reached infrequently are an accurate signal of manager quality. These levels of outperformance mechanically occur more often during volatile times, where a fund may have a better chance to reveal their gambling nature by getting excessively high volatility in the volatility sort, or by getting excessively (un)lucky in the return sort.

In the first attempt to find gambling funds I sort funds into buckets based on their realized, annualized, total volatility in the prior year (t-12 to t). The first bucket is funds with less than 10% volatility. Subsequent buckets are formed in steps of 10% starting with 10% - 20% until 50% is reached. The final bucket is all funds with volatility exceeding 50% in the previous year. Summary statistics for volatility buckets is presented in table (6).

Table 6

Volatility buckets

Annualized σ	Fund-month count	Unique dates represented
$\sigma_i \leq 10\%$	514,039	251
$10\% \leq \sigma_i \leq 20\%$	94,419	251
$20\% \leq \sigma_i \leq 30\%$	4,823	248
$30\% \leq \sigma_i \leq 40\%$	693	171
$40\% \leq \sigma_i \leq 50\%$	171	84
$\sigma_i \geq 50\%$	140	81

Table 6 displays the number of fund-month observations and unique dates represented within each bucket created by sorting funds on their past 12 month volatility.

The fund-month count is the total number of return observations in the bucket across the sample period. The unique dates represented record the number of unique dates where at least one fund is in that bucket. The number of dates represented decreases as buckets get more extreme from the original total number of unique dates for the study, 251. The number of observations also decreases in more extreme buckets. To visually demonstrate the distribution of unique dates in the most extreme volatility bucket, Figure 4 shows the count of observations at each date where at least one fund had annualized volatility in the prior year greater than 50%.

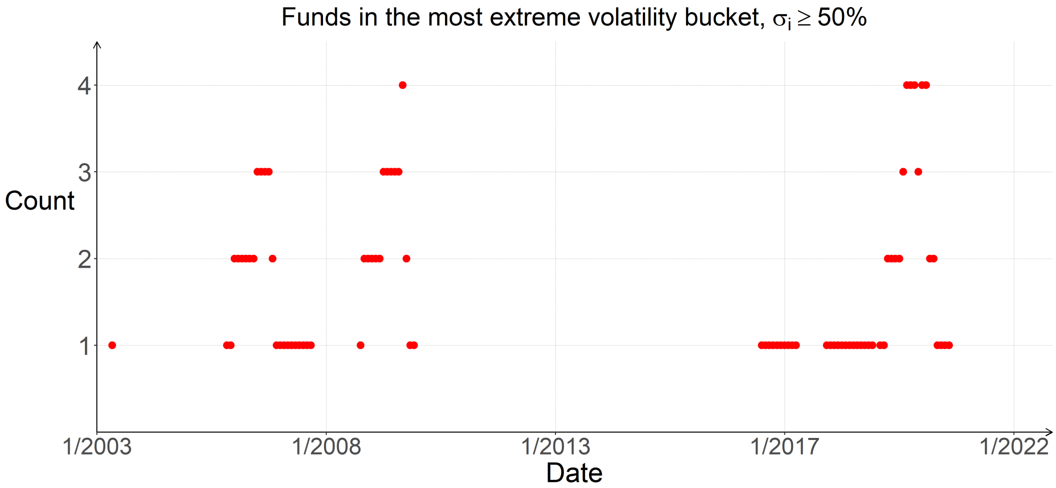

Figure 4. This figure demonstrates the number of funds that appear in the most extreme volatility bucket, with more than 50% annualized volatility in the previous 12 months.

Figure 4 shows the number of funds in the most extreme volatility bucket over time. Funds do not appear in this bucket at every date, and at most, four funds are in this bucket during any given month. Funds in this bucket occur during times of higher market volatility, and tend to occur around the Great Recession as well as during COVID. Time fixed effects will be included in regressions of future performance to ensure that, for example, results are not driven by all mutual funds performing poorly in periods following high market volatility.

In order to investigate the performance of each volatility bucket, I run a regression with the fund's volatility bucket as a dummy variable:

$$y_{i,t} = \alpha + \sum_{k=1}^{5} \beta_k \mathbb{1}_{(a_j < \sigma_{i,t} < b_j)} + Controls_{i,t} + timeFE_t + \epsilon_{i,t} \tag{11}$$

To avoid perfect multicollinearity the baseline bucket (funds with volatility less than 10%) is dropped from the regression. As Table (6) shows, most fund-month observations are in this bucket. The performance of other buckets is measured in comparison with these funds. $y_{i,t}$ will be measured as the cumulative excess return, Sharpe Ratio, and average monthly alpha of funds over the next year.

One problem that arises with monthly sorting and annual performance analysis is that funds have overlapping periods of return. For example, a fund that is sorted into the lowest volatility bucket for two months in a row will have (i) performance from t+1 to t+12 and (ii) performance from t+2 to t+13 represented in that bucket. This introduces mechanical correlation in results within each bucket, which understates standard errors in the typical regression framework. To correct for this, I use standard errors from Thompson (2011) that double cluster on the fund-time level and control for persistent common shocks. The formula for these standard errors are:

$$\hat{Var}(\hat{\beta}) = \hat{V}_{fund} + \hat{V}_{time,0} - \hat{V}_{white,0} + \sum_{l=1}^{L}(\hat{V}_{time,l} + \hat{V}'_{time,l}) - \sum_{l=1}^{L}(\hat{V}_{white,l} + \hat{V}'_{white,l}) \quad (12)$$

$\hat{V}_{fund}$ is the usual formula for standard errors clustered by fund, $\hat{V}_{time,l}$ is the usual formula for standard errors clustered by time, and $\hat{V}_{white,l}$ are the usual OLS standard errors robust to heteroskedasticty. L is the time period of persistent common shocks. With one year horizons, $L = 12$, and for other horizons discussed in the robustness sections L is equal to the number of months performance is projected. In addition to double clustering on the time and fund dimensions, the formula adds an estimation of the weighted autocovariance between time clusters to the usual standard errors, accounting for the mechanical persistence of overlapping return periods. All regressions in the results and robustness sections utilize this formula for standard errors.

Table (7) reports fund performance by volatility bucket with next year's cumulative excess return as the y variable. All regressions in this paper report results as whole percents: for example, the beta of 0.59 in the first column of the the 10 to 20 percent bucket means funds in these buckets on average outperform the omitted group by 0.59% in cumulative excess return over a one year horizon. Stars represent statistical significance at the usual thresholds of 10%, 5%, and 1%.

The original hypothesis states that poor quality managers have higher volatility than good quality managers, predicting that higher volatility is correlated with poor future performance. The first regression conforms to the hypothesis. Funds with higher volatility than the omitted group of funds with less than 10% volatility have monotonically worse cumulative excess return in the year following sorting. The economic significance in funds with extremely high prior year volatility is large: funds with over 50% volatility in the previous year underperform their peers in cumulative excess return by a staggering 23% in the following 12 months. With time fixed effects, the result is not driven by the highest volatility group only appearing in certain months. Results are robust to the inclusion of extra controls, even though about 120,000 observations are dropped for not having expense or turnover ratio information.

Although the first regressions are indicative of poor performance of high volatility funds, it may not be conclusive of overall poor performance. First, it is already

Table 7
Cumulative excess return over the following year by volatility bucket

	(1)	(2)
$10\% \leq \sigma_i \leq 20\%$	0.59	0.87
$20\% \leq \sigma_i \leq 30\%$	-1.43	-1.22
$30\% \leq \sigma_i \leq 40\%$	-9.89***	-7.85***
$40\% \leq \sigma_i \leq 50\%$	-18.48***	-17.46*
$\sigma_i \geq 50\%$	-23.13***	-23.93**
Age	0.77***	0.47***
Size	-0.06	0.02
Expense Ratio		0.71***
Turnover Ratio		-0.57***
N	621,268	495,148
Time FE	yes	yes
Thompson standard errors	yes	yes

Table 7 displays the beta of each volatility bucket from a regression of cumulative excess return in the 12 months following bucket formation on volatility buckets and controls. Stars indicate statistical significance at 10%, 5%, and 1% thresholds for one, two and three stars, respectively. While both columns have age and size control variables, the second column adds expense and turnover ratio variables as controls.

documented that stocks with high idiosyncratic volatility underperform their peers, although the magnitudes found here appear higher than what was documented by papers in that literature (see, e.g., Ang et al. (2006)). Second and more crucially, the excess return is not adjusting for risk; it assumes that all funds have the S&P 500 as their benchmark. Since date fixed effects are used this effectively means funds are measured relative to all other funds, as each funds' return within any given month is linearly shifted downward by the same return on the S&P 500.

Before addressing benchmark adjustments, one possible story to rule out is that results are driven by short-term reversion (Jegadeesh (1990)). It could be the case that high volatility funds are often in the right tail of the return distribution when they are sorted into volatility buckets, and these funds do very poorly over the next few months after sorting. It would be concerning for the hypothesis which claims high volatility managers are poorly skilled if the entire negative return was realized in the first few months after sorting. Figure 5 investigates this explanation by plotting the average cumulative return of funds in the most extreme volatility bucket over the following year. It shows that these funds consistently underperform the market throughout the entire year following sorting, alleviating concerns that the result is driven by short-term reversion.

Although the results on performance in excess of the market conform to the hypothesis and are not driven by short-term reversal, they are based on the assumption that all funds share the same benchmark - the S&P 500. To further evaluate performance, I examine regressions which use the funds' volatility bucket to predict (i)

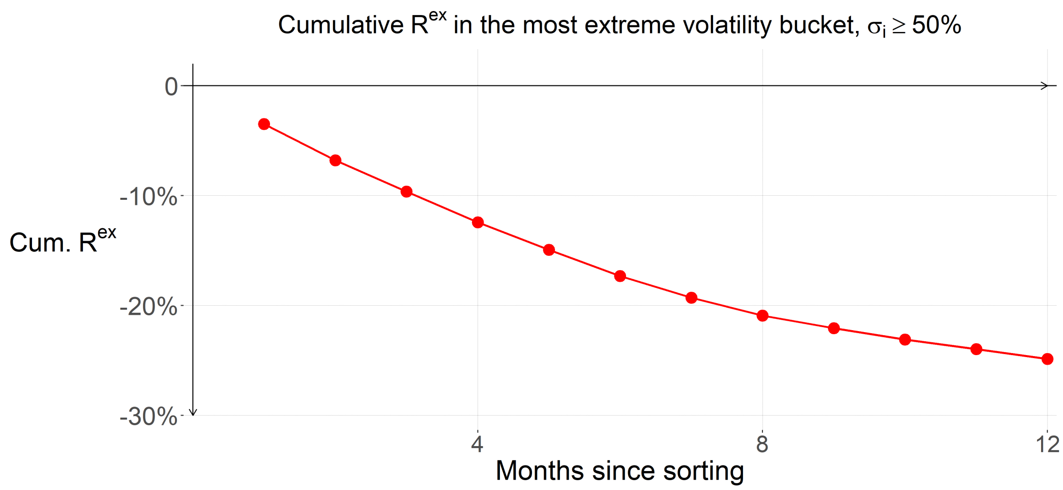

Figure 5. Figure 5 shows the average cumulative excess return of funds in the most extreme volatility bucket (with annualized volatility greater than 50% in the previous year) over the 12 months following bucket formation

Sharpe Ratio and (ii) three and seven factor alphas. Although Sharpe Ratio has been found to be a problematic measure of performance (Lehmann and Timmermann (2007), Lo (2002)) it is appealing in this context because it only requires a fund's return information (meaning no observations are dropped by requiring information on an additional variable) and it carries an inherent penalty for high volatility.

Table 8

Annualized Sharpe Ratio over the following year by volatility bucket

	(1)	(2)
$10\% \leq \sigma_i \leq 20\%$	-0.097***	-0.077*
$20\% \leq \sigma_i \leq 30\%$	-0.470***	-0.403***
$30\% \leq \sigma_i \leq 40\%$	-1.107***	-1.03***
$40\% \leq \sigma_i \leq 50\%$	-1.793***	-1.793***
$\sigma_i \geq 50\%$	-1.415***	-1.395*
Age	0.006	0.008
Size	0.008	0.003
Expense Ratio		-0.053**
Turnover Ratio		-0.034***
N	617,808	492,681
Time FE	yes	yes
Thompson standard errors	yes	yes

Table 8 displays the beta of each volatility bucket from a regression of Sharpe Ratio in the 12 months following bucket formation on volatility buckets and controls. Stars indicate statistical significance at 10%, 5%, and 1% thresholds for one, two and three stars, respectively. While both columns have age and size control variables, the second column adds expense and turnover ratio variables as controls.

The funds' performance in annualized Sharpe Ratio in the year following their sorting shows a similar pattern to excess returns. Funds' performance declines nearly monotonically in past year volatility. However, the most extreme group ($\sigma_i \geq 50\%$)

reverses the pattern and slightly outperforms funds with annualized volatility between 40% and 50% over the previous year. This pattern holds for both the baseline regression and with the inclusion of additional controls. In the specification with expense and turnover ratios the statistical significance of the beta being different from zero in the most extreme group decreases slightly from losses of observations, although the estimate of beta is nearly identical. The number of observations in this bucket drops from 140 to 97 once funds without turnover and expense ratio information are dropped, increasing the standard error of the beta estimate.

The magnitudes of underperformance of the extreme volatility group reported in Table 8 are again staggeringly high; for context, from 2015 to 2025, the S&P 500 had a Sharpe Ratio around 0.71.[5] However, given the funds' extremely low performance in excess of the market and their high volatility, it's unsurprising they have negative and extreme value of Sharpe Ratio.

The most important risk adjustment is in the funds' alpha. By observing the funds' factor exposures through their holdings observations, I can control for fund style to see if they are outperforming their risk exposures. Monthly alphas are calculated from stock betas obtained in quarterly holdings observations as detailed in the data section. The next two regressions in Tables (9) and (10) report average monthly three and seven factor alphas for funds in each bucket in the year following bucket formation.

Table 9

Average monthly 3F alpha over the following year by volatility bucket

	(1)	(2)
$10\% \leq \sigma_i \leq 20\%$	-0.078	-0.058
$20\% \leq \sigma_i \leq 30\%$	-0.320***	-0.286***
$30\% \leq \sigma_i \leq 40\%$	-0.518***	-0.432**
$40\% \leq \sigma_i \leq 50\%$	-0.270	-0.368
$\sigma_i \geq 50\%$	0.557***	0.549***
Age	0.025***	0.024***
Size	-0.007**	-0.011***
Expense Ratio		-0.013
Turnover Ratio		-0.062**
N	256,870	230,765
Time FE	yes	yes
Thompson standard errors	yes	yes

Table 9 displays the beta of each volatility bucket from a regression of 3 factor average monthly alpha for the 12 months following bucket formation on volatility buckets and controls. Stars indicate statistical significance at 10%, 5%, and 1% thresholds for one, two and three stars, respectively. While both columns have age and size control variables, the second column adds expense and turnover ratio variables as controls.

[5] https://www.morningstar.com/indexes/spi/spx/risk

Table 10
Average monthly 7F alpha over the following year by volatility bucket

	(1)	(2)
$10\% \leq \sigma_i \leq 20\%$	0.037	0.046
$20\% \leq \sigma_i \leq 30\%$	-0.119*	-0.110*
$30\% \leq \sigma_i \leq 40\%$	-0.392**	-0.331
$40\% \leq \sigma_i \leq 50\%$	-1.00**	-1.388
$\sigma_i \geq 50\%$	0.351*	0.345*
Age	0.034***	0.033***
Size	-0.005	-0.007**
Expense Ratio		0.002
Turnover Ratio		-0.036
N	256,870	230,765
Time FE	yes	yes
Thompson standard errors	yes	yes

Table 10 displays the beta of each volatility bucket from a regression of 7 factor average monthly alpha for the 12 months following bucket formation on volatility buckets and controls. Stars indicate statistical significance at 10%, 5%, and 1% thresholds for one, two and three stars, respectively. While both columns have age and size control variables, the second column adds expense and turnover ratio variables as controls.

Unlike the results for excess return and Sharpe Ratio, the results for alpha appear to reject the initial hypothesis. Although volatility is initially associated with decreased alpha, above a certain threshold, higher volatility instead improves fund performance. Beginning with the three factor alpha, the trough of fund performance is found in the 30-40% group in both control variable specifications. Funds in this bucket earn a monthly alpha 52 basis points (or 43 basis points with expense and turnover ratios included) less than their comparable peers in the omitted less than 10% bucket. However, in the 40-50% bucket, performance is improved and the negative estimate is no longer statistically significant from zero. In most extreme volatility bucket performance is positive and statistically significant, where funds earn a positive monthly alpha of about 55 basis points per month with either set of controls.

In addition to MKT, SMB, and HML, the seven factor alpha adds UMD (momentum), RMW (profitability), CMA (investment), and STR (short-term reversal factors). The STR factor is chosen to ensure alpha results are not driven by mechanical short term reversal of security prices. The seven factor alpha initially shows the same pattern as the three factor alpha; performance decreases monotonically with volatility until a trough is reached, in this case, in the 40-50% volatility bucket. However, similar to the three factor alpha, the trend reverses in the most extreme bucket, where a positive and (albeit marginally) statistically significant positive alpha are found. This pattern holds across both control variable specifications. In contrast to the results on return and Sharpe Ratio, the statistical significance is low on all alpha results, but especially in the specification with expense and turnover ratio variables added as controls.

The alpha results are in favor of rejecting the initial hypothesis, that bad managers gamble. The initial hypothesis predicts that performance would get monotonically worse, even in the extreme right tail of volatility. Even if the trend reverses, it should never be the case that good managers are found in the extreme tail; the estimate should always be negative, even if it is not statistically significant.

It could be the case that this pattern is caused by statistical biases. A large number of observations were dropped from going from returns to alpha, as funds need to be matched with holdings from Reuters. The total fund-month observations have decreased from 659,734 to 256,870, even before expense and turnover ratio are included as controls. The decrease in observations is especially severe in the tails. Fund-month observations in the 40% to 50% bucket decreases from 171 to 36, while fund-month observations in the $\sigma_i \geq 50\%$ bucket deceases from 140 to 32.

In the introduction I explain that finding beta exposures from holdings has benefits in allowing me to more closely track funds' potentially volatile factor exposures, but that it introduces selection bias: funds that match with holdings from Reuters are on average larger and perform much better compared to funds where only returns data is available. Results should be considered an upper bound on fund performance. It's very likely that the roughly 20% of observations remaining in this bucket after requiring high quality holdings information have been positively selected. Table 11 shows that the remaining funds in this bucket do appear to be very positively selected by regressing cumulative excess return on a dummy for holdings information. This regression is only run on funds in the greater than 50% volatility bucket: the reported beta is the difference in cumulative excess return between funds in this bucket with holdings information compared to funds in this bucket without holdings information. That is, I run:

$$y_{i,t} = \alpha + \beta_1 \mathbb{1}^{Holdings} + \epsilon_{i,t} \tag{13}$$

On the subset of funds with prior year annualized volatility greater than 50%. Note that time fixed effects have not been included as there are only 32 observations with holdings information occurring at different points in time. Additionally, note that statistical significance is only reported for the beta estimate, not the regression intercept.

Table 11 documents that funds in the most extreme volatility bucket with holdings information appear to be performing much better than other funds in the bucket without holdings information. The average cumulative excess return of funds without holdings information in the bucket is about -32%. However, the funds with holdings information in that bucket on average have a cumulative excess return 32% higher: these funds perform roughly the same as the market in the year following

Table 11
Selection in holdings

	Intercept (α)	β_1
$R^{ex}_{t\in 1,12}$	-31.744	32.100***

Table 11 compares the cumulative twelve month excess return in the most extreme volatility bucket. In the first column, the regression intercept is calculated for return of all funds in the bucket without holdings information. In the second column, the beta on the dummy represents the return of funds in this bucket with holdings information is shown.

bucket formation. The fact that these funds with holdings information have positive alpha is unlikely to have high external validity for interpreting the performance of all funds in this bucket. Regardless, without information on the funds without a high quality match to holdings, I cannot conclude that funds in this bucket have statistically significant negative performance as measured by alpha.

One fundamental problem of sorting on volatility is that it does not allow for future performance of fund managers to vary depending on which tail of return the fund was sorted into. It could be the case that funds in the left tail of return (but the right tail of volatility) are affected by survivor bias of a different magnitude, or have different fundamental future performance than right tail return funds. It has been shown, for example, that funds often fire their managers or change their strategy after realizing a left tail return (Lynch and Musto (2003)). The next section of results will sort funds by their excess return rather than volatility to explore this potential heterogeneity in fund performance.

1.3.2 Sorting by returns

This section alters the sorting method to allow for funds' behavior to be different between the left and right tails of the return distribution. Funds in both the left and right tail of the return distribution are mechanically likely to be in the right tail of the volatility distribution, masking potential differences in their future performance with the volatility sort. The baseline bucket will be funds that earn between -2% and 2% return relative to the S&P 500 in any given month. Funds are then sorted into 2-5%, 5-10%, 10-15%, 15%+ buckets in the right tail. Funds with left tail returns are symmetrically sorted into buckets with negative percentages, with the most extreme group being funds that underperform the S&P 500 by more than 15% within any given month. The results in this section will have the same strucure as the section on volatility: the funds' cumulative excess return by bucket is examined, followed by risk adjustments with regressions using Sharpe Ratio and three and seven factor alphas as y variables.

Table 12 summarizes unique dates and counts of funds in each of the return buckets. Similar to the volatility sort, the majority of funds all lie within one baseline group: in this case, funds that earn between -2% and 2% excess return

Table 12
Return buckets

R^{ex}	Fund-month count	Unique dates represented
$R^{ex} < -15\%$	246	61
$-15\% \leq R^{ex} < -10\%$	1,115	125
$-10\% \leq R^{ex} < -5\%$	12,193	240
$-5\% \leq R^{ex} < -2\%$	77,785	251
$-2\% \leq R^{ex} < 2\%$	486,866	251
$2\% \leq R^{ex} < 5\%$	69,797	251
$5\% \leq R^{ex} < 10\%$	10,848	231
$10\% \leq R^{ex} < 15\%$	710	103
$R^{ex} \geq 15\%$	174	60

Table 12 displays the number of fund-month observations and unique dates represented within each bucket created by sorting funds on their contemporaneous pre-fee excess return.

relative to the S&P 500 within the month of their observation. The amount of fund-month observations, and the number of unique dates that they represent, declines monotonically towards either tail of return distributions. To see how the unique dates are distributed, figure 6 plots the counts of funds in the extreme left and right tail of the returns distribution at each date.

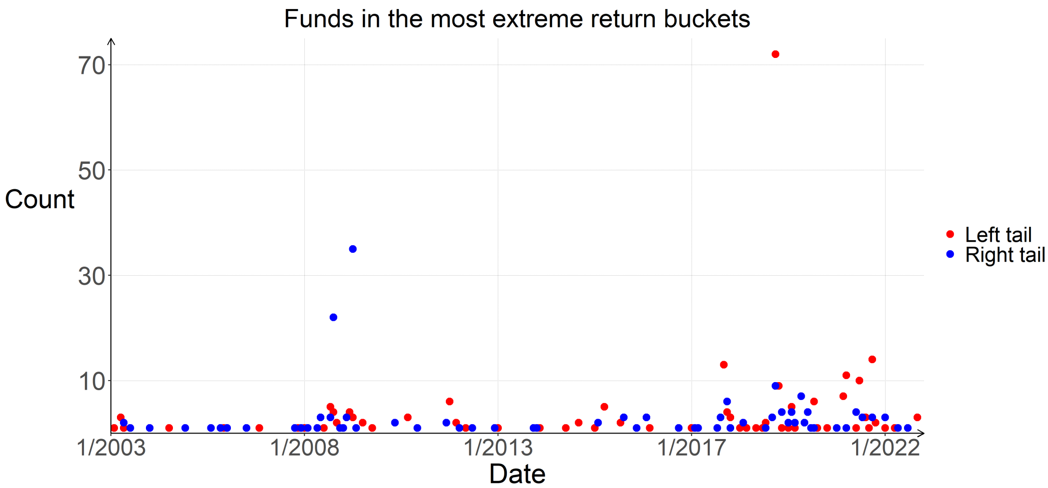

Figure 6. This figure shows the number of funds that appear in the most extreme return buckets in every month. The number of funds in the extreme left tail ($R^{ex} < -15\%$) are denoted in red, while the number of funds in the extreme right tail ($R^{ex} \geq 15\%$) are denoted in blue.

Membership in the most extreme left and right tail buckets are more clustered compared to the volatility sort. The highest number of funds in the extreme right tail occurs in April 2009, when the bear market was beginning to recover during the Great Recession, while the highest number of funds in the extreme left tail is found in March 2020, during COVID.

The results of this section will find that funds in the right tail of the return distribution underperform across multiple measures over one-year horizons[6]. Before continuing on to evaluate the performance of funds in different return buckets, I will briefly summarize the observable characteristics of right tail funds. Table 13 presents results.

Table 13

Summary statistics - right tail funds

	Mean	Median	Standard Deviation	N
Total observations	174	-	-	-
Number of distinct funds	105	-	-	-
TNA (millions)	518.66	117.45	1512.16	174
Age (years)	13.06	11.00	8.98	174
Expense Ratio (%)	1.70	1.55	0.71	154
Turnover Ratio (%)	2.00	1.02	2.84	140
Percent cash (%)	3.10	0	7.00	174

Table 13 displays summary statistics of variables for funds only in the right tail of the short-term return distribution (with contemporaneous pre-fee excess return greater than 15%).

The main purpose of this table is to show that right tail fund results are not driven by exceptionally small or young funds. Funds below 5m TNA have already been removed from the sample, but the average size of extreme right tail funds is well above that cutoff, averaging a size of about 518 million. That is below the mean size of all active funds (about 1,575 million), but well above the 5m cutoff for extremely small funds. Similar to the overall sample the median is well below the mean, as the size distribution of right tail funds is influenced by the existence of a few very large funds. The median of right tail funds, 117.45m, is about half the overall sample median size of 262.80m.

In contrast to size, the age distribution of right tail funds is roughly in line with the age distribution of the entire sample. Right tail funds have a slightly lower mean age of 13.06 (compared with 13.58 years in the main sample) but a slightly higher median age of 11.00 (compared with 10.00 in the full sample).

Funds in the right tail charge higher fees and have a higher turnover ratio than the full-sample averages. The turnover ratio will be investigated in much more detail in the section of results where I investigate gambling behavior. They have higher mean cash than the full sample, although the median fund has no data on cash holdings. Cash holdings were set to zero for funds missing this variable, meaning beta exposures are not adjusted by cash holdings when computing alpha. Note that alpha results are robust to adjusting or not adjusting beta exposures by cash.

As with the volatility sort, analysis of performance will begin with the funds' excess return in Table 14. Identically to the previous section, all regressions include date fixed effects and utilize Thompson (2011) standard errors to double cluster and control for persistent common shocks. The full regression equation is denoted as

[6] The robustness section discusses other performance horizons.

$$y_{i,t} = \alpha + \sum_{k=1}^{8} \beta_k \mathbb{1}_{(a_j < R^{ex}_{i,t} < b_j)} + Controls_{i,t} + timeFE_t + \epsilon_{i,t} \tag{14}$$

Where $y_{i,t}$ initially takes on the value of funds' cumulative excess returns in the 12 months following bucket formation. The -2% to 2% bucket is omitted as a baseline.

Table 14

Cumulative excess return over the following year by return bucket

	(1)	(2)
$R^{ex} < -15\%$	-0.228	2.870
$-15\% \leq R^{ex} < -10\%$	-2.15	-2.273
$-10\% \leq R^{ex} < -5\%$	-2.282**	-2.071**
$-5\% \leq R^{ex} < -2\%$	-0.291	-0.235
$2\% \leq R^{ex} < 5\%$	0.696	0.720
$5\% \leq R^{ex} < 10\%$	0.735	0.805
$10\% \leq R^{ex} < 15\%$	0.791	0.022
$R^{ex} \geq 15\%$	-10.134**	-8.202
Age	0.6727***	0.397**
Size	-0.047	0.031
Expense Ratio		0.752**
Turnover Ratio		-0.516***
N	621,268	495,148
Time FE	yes	yes
Thompson standard errors	yes	yes

Table 14 displays the beta of each return bucket from a regression of cumulative excess return in the 12 months following bucket formation on return buckets and controls. Stars indicate statistical significance at 10%, 5%, and 1% thresholds for one, two and three stars, respectively. While both columns have age and size control variables, the second column adds expense and turnover ratio variables as controls.

In contrast to the results with the volatility sort, fund performance does not monotonically decrease towards both tails. In the left tail, performance initially declines to a trough in the -5% to -10% bucket. Performance stays roughly constant when moving to the -10% to -15% bucket, but reverses when the extreme left tail is reached. In the right tail, funds initially outperform their peers in market-adjusted return over the following year. Although not statistically significant, the beta estimate is positive for the 2-5%, 5-10%, and 10-15% groups. However, the extreme right tail is clearly special; funds earning more than 15% than the market in the current month underperform other mutual funds by 10.1% over the following year in the specification without expense and turnover ratios. This result is statistically significant at the 5% level. The estimate is slightly improved but still very negative when expense and turnover ratios are included as controls. However, in this specification, the result is not statistically significant. Remember at the end of the Data section I documented that funds with non-missing expense and turnover ratios are positively selected relative to the full sample of funds. This positive selection and a decrease in observations from 174 (full sample) to 139 (requiring expense and turnover ratios) in the most extreme bucket is enough to push the p-value to about 12%, just outside the first threshold of statistical significance.

Comparing this result to the volatility sort, it's clear that funds perform differently in the extreme left and right tails of the distribution. Although both would have mechanically high volatility by achieving an extreme return, the extreme left tail outperforms the funds in other negative return buckets while the extreme right tail underperforms other buckets in the right tail. There is evidently something special about the funds sorted into either tail. If both of these ended up in high volatility groups, it is right tail funds that are driving negative results reported in the previous section.

Another contrast between the return volatility sorts is that the negative monotonicity of performance in both tails has disappeared. A negative pattern is still somewhat evident in the left tail, where the beta of next year's cumulative excess return declines from close to zero in the -5% to -2% group to about -2% in the -10% to -5% and -15% to -10% groups. However, in the right tail, funds seem to outperform the market until the extreme right tail is reached. Although not statistically significant, beta estimates are slightly improving in the moderate right tail, from +0.69% in the 2% to 5% group to +0.79% in the 10 to 15% group. This pattern suggests a model slightly different than what is graphed in the introduction (Figure 1) - the result on cumulative excess return suggests more segmented return possibilities than illustrated by the figure, where moving towards the right tail more smoothly increases your probability of finding a bad manager.

Even more so than with the volatility sort, short-term reversal is a concern to the interpretation of results when funds are sorted by excess return. If extreme right tail funds have an excess return of, for example, -10% in the first month following bucket formation their year-long cumulative excess return being negative is no longer an indicator of the manager being poor quality. A story of results driven by short-term reversal is not fully consistent with the results presented in Table 14, as the left tail would then be outperforming the market and the moderate left tail would be underperforming the market. However, to alleviate concerns that the extreme right tail result may be driven by short-term reversal, Figure 7 plots the average cumulative return of funds in the most extreme return bucket over the following year.

Figure 7 alleviates the concern that negative performance in the right tail is driven by mechanical short-term reversal. The funds initially perform roughly the same as the market in the first two months but then outperform the market slightly by month five. This could possibly be driven by flows into the funds providing upwards pressure to the stock prices of their holdings (Lou (2012)), or simply by

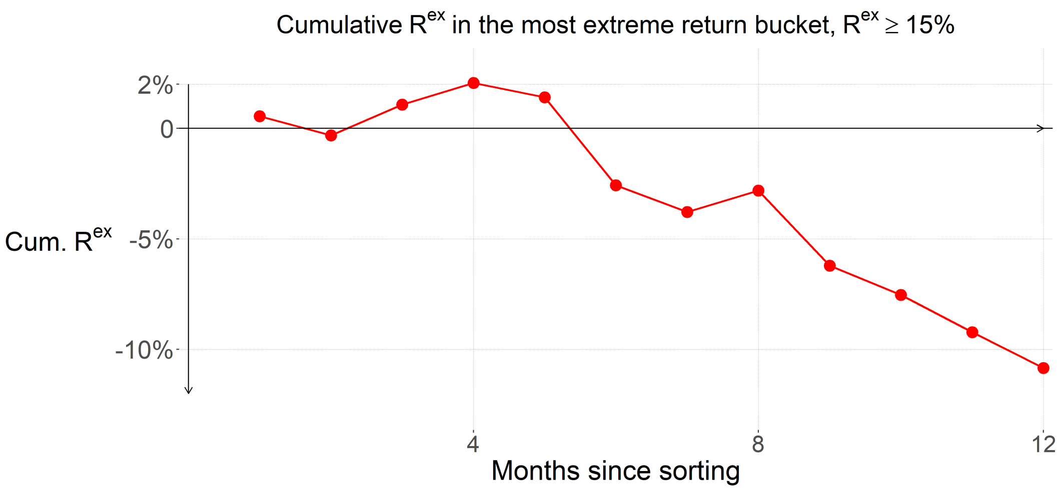

Figure 7. Figure 7 shows the average cumulative excess return of funds in the most extreme right tail return bucket (funds that beat the market by more than 15% in the current month) over the 12 months following bucket formation.

momentum, which has peak returns in 6-12 month horizons (Jegadeesh and Titman (2001)). In the sixth month since bucket formation funds on average underperform the market by over 4%. Between six and twelve months right tail funds have underperformed the market cumulatively by more than 10%. The poor performance of right tail funds over the next year is driven by consistent negative performance in the six month to one year horizon. In the robustness section, horizons other than one year are considered; the extreme right tail additionally performs poorly on cumulative excess return on two, three, and five year horizons. The only horizon where the extreme right tail bucket's performance is not statistically significant is when performance is examined over six months, although the beta is still negative. The negative performance result strengthening over longer horizons is the opposite of what we should find if results were driven by mechanical short-term reversion.

Evaluation of funds' risk-adjusted performance will begin with an analysis of funds' annualized Sharpe Ratio in the 12 months following bucket formation. Regression results are presented in Table 15.

The Sharpe Ratio results differ from the excess return results in the middle of the distribution and the moderate tails. When performance is penalized for high volatility, the omitted group with excess return -2 to 2% is the highest performing bucket. Fund performance monotonically declines towards both tails until the extreme buckets are reached. Additionally, the extreme left tail now underperforms all other left tail buckets. Similar to the excess return results, the extreme right tail underperforms all other buckets.

The most important risk adjustment for performance is when alpha is considered as the dependent variable in the regression. Tables 16 and 17 report average monthly alphas with three and seven factor beta specifications.

Beginning with Table 16, monthly three factor alphas monotonically decline towards the extreme left tail in both specifications for control variables. However,

Table 15
Annualized Sharpe Ratio over the following year by return bucket

	(1)	(2)
$R^{ex} < -15\%$	-0.723***	-0.628***
$-15\% \leq R^{ex} < -10\%$	-0.467***	-0.446***
$-10\% \leq R^{ex} < -5\%$	-0.233***	-0.214***
$-5\% \leq R^{ex} < -2\%$	-0.091***	-0.085***
$2\% \leq R^{ex} < 5\%$	-0.525*	-0.042
$5\% \leq R^{ex} < 10\%$	-0.155**	-0.143**
$10\% \leq R^{ex} < 15\%$	-0.541***	-0.493***
$R^{ex} \geq 15\%$	-1.060***	-0.913***
Age	0.004	0.006
Size	0.009*	0.003
Expense Ratio		-0.058**
Turnover Ratio		-0.037***
N	617,808	492,681
Time FE	yes	yes
Thompson standard errors	yes	yes

Table 15 displays the beta of each return bucket from a regression of Sharpe Ratio in the 12 months following bucket formation on return buckets and controls. Stars indicate statistical significance at 10%, 5%, and 1% thresholds for one, two and three stars, respectively. While both columns have age and size control variables, the second column adds expense and turnover ratio variables as controls.

once the extreme left tail is reached, fund performance estimates turn positive, albeit without statistical significance. In the right tail performance monotonically declines towards the extreme right tail mostly without any level of statistical significance. This pattern continues to the extreme right tail, where funds earn a monthly alpha 29 basis points worse than the omitted group in the first specification without expense and turnover ratios. In the specification with expense and turnover ratios, similar to the results on excess return, the loss of observations and positive bias of including extra variables pushes the estimate above the 10% threshold with the robust standard errors. The p-value for this estimate is just below 14%.

In the seven factor alpha presented in Table 17, extra factors explain much of the negative return in the moderate right tail. The mid-left tail buckets still underperform the omitted group, but by less extreme values and mostly without statistical significance. The mid-right tail of return has alphas essentially equal to zero - they are no longer underperforming the omitted group. However, the extreme left and right tails are still outliers of performance. The extreme left tail has positive estimates of alpha. In the specification with expense and turnover ratios, the positive alpha is statistically significant at the 10% level. The extreme right tail continues to perform poorly. The magnitudes in each specification are strikingly similar to the magnitudes of the three factor alpha in Table 16. The additional factors did reduce the variance of alpha estimates, however, making the negative estimate statistically

Table 16
Average monthly 3F alpha over the following year by return bucket

	(1)	(2)
$R^{ex} < -15\%$	0.046	0.083
$-15\% \leq R^{ex} < -10\%$	-0.33***	-0.304***
$-10\% \leq R^{ex} < -5\%$	-0.169***	-0.158***
$-5\% \leq R^{ex} < -2\%$	-0.053**	-0.046**
$2\% \leq R^{ex} < 5\%$	-0.006	0.029
$5\% \leq R^{ex} < 10\%$	-0.115*	-0.090*
$10\% \leq R^{ex} < 15\%$	-0.165	-0.149
$R^{ex} \geq 15\%$	-0.292**	-0.236
Age	0.025***	0.025***
Size	-0.008**	-0.013***
Expense Ratio		-0.018
Turnover Ratio		-0.062**
N	256,870	230,765
Time FE	yes	yes
Thompson standard errors	yes	yes

Table 16 displays the beta of each return bucket from a regression of 3 factor average monthly alpha for the 12 months following bucket formation on return buckets and controls. Stars indicate statistical significance at 10%, 5%, and 1% thresholds for one, two and three stars, respectively. While both columns have age and size control variables, the second column adds expense and turnover ratio variables as controls.

significant at the 1% level in the initial specification and significant at the 5% level in the specification with the expense and turnover ratio variables added as controls.

Two clear patterns emerge from the investigation of future fund performance from the return sort. First, the extreme left tail performs surprisingly well. It outperforms the mid-left tail in excess return and both measures of alpha, and only underperforms these buckets when Sharpe Ratio is considered. The most robust pattern across all measures of performance is the underperformance of the extreme right tail. In excess return, Sharpe Ratio, and three and seven factor alphas, funds that beat the market by more than 15% within their month do very poorly in the following year. Measures of performance for these funds are always negative and almost always significant, with a few exceptions due to p-values rising a little above 10% when expense and turnover ratios are included as controls.

One final measure of performance is considered by looking at flows into these funds in the month following their return realization, and looking at average flows out of these funds in the year following their return realization. By measuring average flows over time I can observe if the average right tail fund is pursuing a strategy that is supported by their investors.

Table 17
Average monthly 7F alpha over the following year by return bucket

	(1)	(2)
$R^{ex} < -15\%$	0.179	0.215*
$-15\% \leq R^{ex} < -10\%$	-0.202***	-0.183***
$-10\% \leq R^{ex} < -5\%$	-0.031	-0.030
$-5\% \leq R^{ex} < -2\%$	-0.034*	-0.035*
$2\% \leq R^{ex} < 5\%$	0.04	0.041
$5\% \leq R^{ex} < 10\%$	-0.002	0.004
$10\% \leq R^{ex} < 15\%$	0.027	0.017
$R^{ex} \geq 15\%$	-0.237***	-0.232**
Age	0.035***	0.034***
Size	-0.007*	-0.008***
Expense Ratio		0.007
Turnover Ratio		-0.033
N	256,870	230,765
Time FE	yes	yes
Thompson standard errors	yes	yes

Table 17 displays the beta of each return bucket from a regression of 7 factor average monthly alpha for the 12 months following bucket formation on return buckets and controls. Stars indicate statistical significance at 10%, 5%, and 1% thresholds for one, two and three stars, respectively. While both columns have age and size control variables, the second column adds expense and turnover ratio variables as controls.

1.3.3 Flow and flow reversal

Before concluding the results section of the return sort, I examine flows into and out of these funds at one month and one year horizons. It could be the case that despite underperforming in all measures consumers like right tail funds that gamble because they sometimes get spectacularly high returns. In other words, this section aims to discover if funds are behaving as their investors would like them to. If investors support the funds' strategies, they would not withdraw their money from the fund even as it underperformed the market by large margins in the year following bucket formation. Tables 18 and 19 show results of regressions run with one and twelve month flow calculations as the dependent variable. Note that in the second table the number displayed is the average monthly flow for the next twelve months, not the cumulative twelve month flow.

At a one month horizon I reproduce results found in many other papers of the convex flow-performance relationship in mutual funds (see, e.g., Chevalier and Ellison (1997)). Money flows into funds with positive performance and out of funds with negative performance. Left tail funds are not punished as severely as right tail funds are rewarded: in first specification of control variables, for example, extreme right tail funds receive 3.62% inflows while extreme left tail funds only suffer 1.83% outflows.

Table 18
One month flow response by return bucket

	(1)	(2)
$R^{ex} < -15\%$	-1.830***	-1.445***
$-15\% \leq R^{ex} < -10\%$	-0.983***	-0.740*
$-10\% \leq R^{ex} < -5\%$	-0.8434***	-0.740***
$-5\% \leq R^{ex} < -2\%$	-0.467***	-0.430***
$2\% \leq R^{ex} < 5\%$	0.470***	0.636***
$5\% \leq R^{ex} < 10\%$	1.171***	1.580***
$10\% \leq R^{ex} < 15\%$	2.902***	3.713***
$R^{ex} \geq 15\%$	3.623**	4.414***
Age	-1.383***	-1.273***
Size	0.047***	-0.028**
Expense Ratio		-0.570***
Turnover Ratio		-0.144***
N	659,611	536,147
Time FE	yes	yes
Thompson standard errors	yes	yes

Table 18 displays the beta of each return bucket from a regression of flow in the month following bucket formation on return buckets and controls. Stars indicate statistical significance at 10%, 5%, and 1% thresholds for one, two and three stars, respectively. While both columns have age and size control variables, the second column adds expense and turnover ratio variables as controls.

The pattern mostly reverses over yearly horizons. Most variables are very unstable between inclusion of controls, and almost none are statistically significant; there is a large spread of average monthly flows between funds even in the same bucket, as can be expected with differing amounts of managerial skill, strategy, and luck between funds over one year horizons. Despite having +3.62% inflows one month after bucket formation, after twelve months, funds in the extreme right tail have on average suffered negative monthly outflows in the year following bucket formation, albeit without statistical significance. The twelve month average does include the initial positive monthly flow, meaning flows after observing performance from months 2-12 are very negative. It does not appear that investors are pleased with the negative performance of extreme right tail funds in the year following bucket formation.

Before examining how funds are gambling in order to get tail realizations of return, I will briefly discuss possible interpretations of the high performance of funds in the extreme left tail.

1.3.4 Performance in the left tail

The positive performance of the extreme left tail is somewhat puzzling. Although almost no results are statistically significant, the extreme left tail is almost always

Table 19

Twelve month average flow response by return bucket

	(1)	(2)
$R^{ex} < -15\%$	3.098	0.782
$-15\% \leq R^{ex} < -10\%$	-1.021	-4.990
$-10\% \leq R^{ex} < -5\%$	-2.245	-5.022*
$-5\% \leq R^{ex} < -2\%$	-2.386	-5.777
$2\% \leq R^{ex} < 5\%$	0.032	-2.498
$5\% \leq R^{ex} < 10\%$	5.014	-3.048
$10\% \leq R^{ex} < 15\%$	-0.139	-1.347
$R^{ex} \geq 15\%$	-0.930	-3.580
Age	-2.509	-1.313*
Size	0.984	0.959
Expense Ratio		2.891
Turnover Ratio		2.267
N	621,286	495,148
Time FE	yes	yes
Thompson standard errors	yes	yes

Table 18 displays the beta of each return bucket from a regression of average flow in the 12 months following bucket formation on return buckets and controls. Stars indicate statistical significance at 10%, 5%, and 1% thresholds for one, two and three stars, respectively. While both columns have age and size control variables, the second column adds expense and turnover ratio variables as controls.

associated with positive betas for performance, beating the omitted group as well as reversing a typically worsening performance trend in the mid-left tail. It is not driven by short-term reversal, as the seven factor alpha explicitly includes short-term reversal as a factor. Additionally, previous research (see, e.g., Cuthbertson et al. (2010)) has shown that the left tail of mutual funds (sorted by long-term alpha instead of short-term return) persistently underperforms other funds on multiple measures of performance, including alpha. Underperformance of mid-left tail funds is supported by my results, but is reversed when the extreme left tail is reached. How is it that these extreme left tail funds are able to reverse their performance?

The first and most obvious explanation to rule out is survivorship bias. When a fund has had such a drastically low return, it may be the case that most of them fail; the only ones who survive and have a statistic for next year's performance are the funds that do incredibly well, possibly by luck, in the several months following the initial realization. Table 20 investigates the possibility by examning how many observations are lost between one month and one year of the extreme left tail compared to other groups.

Table 20 documents survivorship of different funds in the one year horizon. Three groups are considered. First, as a baseline, all funds are considered. Of the 659,734 fund-month return observations from the initial sorting procedure, 586,673 of them have non-missing return observations every month for the following year, meaning

Table 20
Survivorship of funds

	Initial obs.	Obs. (1 year)	Percentage decline
All funds	659,734	586,673	11.07%
$-15\% \leq R^{ex} < -10\%$	1,115	919	17.58%
$R^{ex} < -15\%$	246	209	15.04%

Table 20 displays summary statistics of survivorship in all funds, funds with pre-fee excess return between -15% and -10%, and extreme left tail funds with contemporaneous pre-fee excess return less than -15%. In the first column, total fund-month observations are tallied. The second column records how many funds remain after 12 months, and the third column calculates the percentage decline in funds in the year following bucket formation.

around 11% of funds failed (or were no longer tracked by CRSP) within the year following bucket formation. Two other groups are considered. For comparison, the count of funds in the second most extreme left tail group (funds with -15% to -10% excess return) is considered. These funds have a documented negative performance over the next year; that is, they performed negatively despite any possible survivorship bias. The number of funds that survives for one year following bucket formation declines from 1,115 observations to 919 observations, losing 17.58% in the following year. The extreme left tail lost more than the average fund but less than the -15% to -10% group, with about 15% of funds failing in the following year. Considering the funds in the -15% to -10% group had persistently negative performance with more failures, the positive performance of the extreme left tail does not appear to be driven by survivorship bias alone.

Although performance estimates are upper bounds of performance, especially when expense and turnover ratio variables are used as controls and when betas are required for alpha estimation, this selection did not stop the mid left tail from achieving negative performance. It appears possible that extreme left tail funds have genuinely improved their performance. However, to reiterate, most positive estimates of performance for extreme left tail funds are not statistically significant.

In (Lynch and Musto (2003)), the authors investigated why funds in the left tail did not receive as much outflows as funds in the right tail received for inflows. They found that funds in the left tail of the return distribution often dramatically change their strategy and/or fire their incumbent manager, providing a rationale for keeping your money in the fund; the extreme left tail return is not indicative of future returns with so much change. It appears their strategy turnarounds may be effective medicine for the most extreme left tail. Interpreting this in the context of my original hypothesis, it is an initially surprising but intuitive complement to my results: bad managers who gamble and get unlucky are discovered and fired, while bad managers who gamble and get lucky are able to keep pursuing value-destroying strategies for some time.

The next section investigates if and how these extreme right tail funds are gambling to achieve monthly excess returns in the right and left tail of the return distribution.

1.4 Investigating Gambling Behavior

The initial hypothesis posits that managers in the tail of the return distribution will perform poorly in the future because these managers achieve tail returns through 'gambling'. By increasing the volatility of their returns, these managers sometimes perform spectacularly well but on average will perform poorly. The focus of this section is investigating if right tail funds gamble, and if so, how they gamble.

The section focuses on the funds in the extreme right tail of the return distribution, and looks for evidence of (i) extreme strategies, (ii) transient strategies, and (iii) the funds' behavior just before their tail return realization. If funds take extreme positions in any way (for example, by having high leverage) and combine it with transient strategies (for example, by having high turnover) I consider the fund to be *gambling*: they are not taking extreme long-term positions to follow strategy but are are taking extreme positions *and* changing them rapidly. As a final check I investigate if the funds' strategies were more transient in the holdings observation just before the extreme return was realized, to see if this behavior was especially pronouned before they realized their right tail return.

1.4.1 Extreme strategies

The first part of the gambling investigation tests if funds in the right tail of the return distribution are using extreme strategies. Three methods of taking extreme positions are examined: first, the fund could be concentrating their portfolio among relatively few stocks in order to increase their idiosyncratic volatility. Second, the fund could be highly levered, or third, the fund could be taking extreme positions in factor exposures.

This first part of the analysis focuses on funds' long-term behavior: do the funds in general have extreme strategies? A later section will investigate funds' behavior in the holdings report just before an extreme return was realized. A dummy variable is created which equals to one if a fund is ever in the right tail of the return distribution during the fund's lifetime. Then, for all funds, I find fund level averages of the number of shares, portfolio concentration of shares, and cash positions (leverage) across all holdings observations. A simple regression is carried out with the following specification:

$$y_i = \alpha + \beta_1 \mathbb{1}^{RT} + \epsilon_{i,t} \tag{15}$$

Where $\mathbb{1}^{RT}$ equals one if the fund is ever in the extreme right tail of returns during its lifetime, and $y_{i,t}$ takes the value of average number of shares, average portfolio concentration, and average cash position of funds. As reported in the data section, HHI measures portfolio concentration by computing

$$HHI_{i,t} = \sum_{j=1}^{J} w_j^2 \tag{16}$$

For each stock j in the fund's holdings. HHI is higher in more concentrated portfolios. Table 21 reports β_1 of the regressions. Note that statistical significance is only reported for the beta estimate, not on the regression intercept (α), which represents the mean of non-right tail funds. In the first table, no beta was significant at any threshold, so no stars are shown.

Table 21
Portfolio concentration and leverage

	α	β_1
Number of shares	151.664	4.984
HHI	0.043	0.009
Cash (%)	3.049%	0.511%

Table 21 displays results of regressions of measures of portfolio concentration and leverage on a dummy variable for extreme right tail performers. The first column reports the regression intercept, while the second reports the beta on the right tail dummy. Statistical significance is shown at the usual thresholds: 10%, 5%, and 1% for one, two and three stars, respectively.

Table 21 shows that funds who earn right tail returns, over their lifetime, are not statistically significantly more extreme in any of the initial measures. Right tail funds have slightly higher portfolio concentration as measured by HHI, but the difference is not statistically significant - the p-value is 0.72. However, they typically have *more* total shares and a *larger* cash position than other funds, which is the opposite of what would be expected if the funds were gambling with large positions in a handful of stocks. Mutual fund regulation limits this type of gambling in any case; funds cannot have more than 5% of their portfolio in any one stock[7] and have limits to borrowing cash. Additionally, almost all funds choose to have positive cash well below the leverage regulation in order to facilitate their liquidity transformation services (Chernenko and Sunderam (2016)).

Another way for funds to have an extreme strategy is for them to have extreme beta exposures. As a reminder, if right tail funds are found to take extreme strategies, this is not a final indictment of 'gambling': it is possible and indeed likely that some funds take extreme factor exposures as instructed by their investors. Aside from loading on idiosyncratic risk, concentrating their portfolios, or taking on excessive leverage, extreme factor exposures is another way to mechanically achieve tail returns. First, the absolute value of factor exposures is examined as a left hand side variable: an extreme long or short exposure to a factor can both result in ex-

[7] An exception has been made in recent years for index funds where over 5% of the index is concentrated in a single share.

treme returns. Table 22 runs the same regression as the previous table but with the absolute value of factor exposures as the left hand side variable. As a reminder, this is looking at the funds' behavior long term: the left hand side variable is the average absolute value of the factor exposure of a fund over its lifetime. For simplicity, only seven factor exposures are reported instead of reporting three and seven factor exposures separately.

Table 22

Average absolute seven factor exposures in the right tail

	α	β_1
MKT	1.004	0.100***
$\lvert SMB \rvert$	0.323	0.196***
$\lvert HML \rvert$	0.192	0.106***
$\lvert UMD \rvert$	0.095	0.054***
$\lvert RMW \rvert$	0.179	0.121***
$\lvert CMA \rvert$	0.209	0.079***
$\lvert STR \rvert$	0.0007	0.0005***

Table 22 displays results of regressions of the average absolute value of seven factor exposures on a dummy variable for extreme right tail performers. The first column reports the regression intercept, while the second reports the beta on the right tail dummy. Statistical significance is shown at the usual thresholds: 10%, 5%, and 1% for one, two and three stars, respectively.

For clarity, note that the column headers of α and β_1 are the results of the regression run with factor exposures as the left hand side variable. Also, MKT has not been listed in the absolute value brackets because the mean of this factor is 1 instead of 0.

As Table 22 shows, right tail funds do follow extreme strategies as measured by factor exposures. They are not very heavily exposed to MKT, which intuitively makes sense; these funds *beat* the market by extreme margins, and they don't have high leverage. But in all other factors right tail funds are following more extreme strategies than other funds, with the absolute value of their exposure over their lifetime often around 50% higher than other funds. Their factor exposures, rather than their portfolio concentration or leverage, explains how they were able to obtain right tail returns.

Before moving on to the analysis of how transient the funds' strategies are, I repeat the analysis of Table 22 with beta exposures instead of the absolute value of beta exposures as the left hand side variable. This table aims to investigate if any stylistic pattern can be discerned from right tail funds. Table 23 summarizes the results.

Note that MKT is not reported as the value is identical to the previous table. In contrast to the results with the absolute value of factor exposures, when raw factor exposures are considered, a pattern emerges on four of six factors (excluding MKT). Funds in the right tail tend to be concentrated in small firms with weak profitability.

Table 23
Average seven factor exposures in the right tail

	α	β_1
SMB	0.244	0.246***
HML	0.044	0.004
UMD	-0.052	-0.030***
RMW	0.022	-0.077***
CMA	-0.089	-0.012
STR	0.0003	0.0004***

Table 23 displays results of regressions of average value of seven factor exposures on a dummy variable for extreme right tail performers. The first column reports the regression intercept, while the second reports the beta on the right tail dummy. Note that the MKT factor is not reported as it is identical to the previous table.

They tend to be short on momentum and to be somewhat exposed to mechanical short-term reversal, although the economic significance of their exposure to STR is not high. There is no clear pattern of right tail funds holding value vs. growth stocks or holding firms with conservative vs. aggressive investment strategies.

1.4.2 Transient Strategies

Funds pursuing extreme but persistent strategies over the long run cannot be classified as 'gamblers': they would simply be carrying out an extreme strategy requested by investors. This next section aims to answer how persistent right tail funds are with their strategies.

Variables in this section will once again be at the fund level - that is, I take the average of a certain variable over the funds' entire lifetime, then compare funds that were in the extreme right tail for at least one month against other funds. The regression specification is identical to the section above, with a dummy variable for membership in the right tail:

$$y_i = \alpha + \beta_1 \mathbb{1}^{RT} + \epsilon_{i,t} \qquad (17)$$

The first two variables of interest are (i) the funds' turnover ratios and (ii) their change in HHI. Do right tail funds trade more often or change their portfolio concentration more often than the average fund? Table 24 reports how right tail fund vary from other funds in terms of these variables.

The right tail trades nearly twice as much as the average fund, redundantly trading about 1.65% of their positions every year. The high levels of turnover begin to indicate the funds may not be following coherent long term strategies. They change their portfolio concentration as measured by HHI more often than the average fund but the difference is not economically or statistically significant.

Given the funds achieve extreme returns by taking extreme factor exposures, the most important metrics to check for gambling is how often their fund exposures

Table 24
Turnover and ΔHHI

	α	β_1
Turnover ratio	0.891%	0.746%***
ΔHHI	-0.004	0.008

Table 24 displays results of regressions of measures of rapidly changing performance on a dummy variable for extreme right tail performers. The first column reports the regression intercept, while the second reports the beta on the right tail dummy.

change. The funds may be trading often but keeping their overall strategy and factor exposures similar. Table 25 reports the standard deviation of beta exposures of right tail funds and all other funds.

Table 25
Fund σ of factor exposures in the right tail

	α	β_1
MKT	0.078	0.059***
SMB	0.106	0.063***
HML	0.128	0.082***
UMD	0.073	0.060***
RMW	0.146	0.092***
CMA	0.161	0.106***
STR	0.0007	0.0006***

Table 25 displays results of regressions of the standard deviation of factor exposures on a dummy variable for extreme right tail performers. The first column reports the regression intercept, while the second reports the beta on the right tail dummy. Statistical significance is shown at the usual thresholds: 10%, 5%, and 1% for one, two and three stars, respectively.

Table 25 shows that funds in the right tail change their factor exposures much more rapidly than other funds. The standard deviation of factor exposures in right tail funds is always positive and statistically significant compared to the average fund. Funds do appear to be gambling - not just by trading more frequently than the average fund, but by trading stocks that differ from their portfolio's current factor exposures. This is a novel and interesting dimension of fund behavior revealed by measuring factor exposures from fund holdings, which show changes in strategy more quickly than could be obtained by conventional rolling window regressions of fund returns on factors.

Before moving on to the robustness section, I consider one last aspect of fund behavior. How much did the funds trade, change their portfolio concentration, or change their beta exposures in the holdings observation directly before they received their right tail return?

1.4.3 Fund behavior just before realization

In this section I investigate how quickly funds in the right tail changed their holdings and factor exposures in the holdings observation directly before receiving their right tail return. This section does not offer definitive proof of gambling, but if funds were trading rapidly or changing their exposures even more often than their own long run average, it offers evidence in favor of the interpretation that funds' gambling was responsible for their lucky return.

In this section I return to regressions at the fund-month level instead of the holdings level. Specifically, I run

$$y_{i,t} = \alpha + \sum_{k=1}^{8} \beta_k \mathbb{1}_{(a_j < R^{ex}_{i,t} < b_j)} + Controls_{i,t} + \epsilon_{i,t} \tag{18}$$

With different left hand side variables as measures of gambling taken from the holdings observation prior to bucket formation. The intercept plus the dummy variable for extreme right tail buckets will be compared against the sum of the intercept and right tail dummy in the 'transient strategies' subsection. For convenience, this sum is reported in the fourth column of Table 26. If the sum of coefficients just before they received their extreme positive return is more than their own long-run averages it suggests the gambling was responsible for the funds' suspiciously high performance. Table 26 documents right tail funds' turnover ratio and change in portfolio concentration just before their extreme positive return. For simplicity, only the regression intercept and coefficient for the dummy variable in the extreme right tail are reported, as the long run behavior of funds by group was only examined for extreme right tail funds in the previous subsection.

Table 26

Turnover and ΔHHI in the holdings observation before sorting

	α	β_1	Sum	Comparison: LR average
Turnover ratio	0.155%	0.908%***	1.063%	1.637%
ΔHHI	-0.004	0.008	0.006	0.004

Table 26 displays results of regressions of measures of transient strategies on a dummy variable for extreme right tail performers. The left hand side variable is the change in HHI and turnover ratio in the holdings observation that matches with the fund's contemporaneous excess return greater than 15%. The first column reports the regression intercept, the second reports the beta on the right tail dummy, and the third reports the sum of the coefficients (which is the average of these variables in right tail funds). The fourth column reports the long run average of these variables for all fund, which is the sum of both columns in Table 25. Statistical significance is shown at the usual thresholds: 10%, 5%, and 1% for one, two and three stars, respectively.

Table 26 reports the turnover ratio and change in portfolio concentration of extreme right tail funds in the holdings observation just before the tail return was realized. The third column, reporting the sum of coefficients, should be compared against the long-run average turnover and change of HHI in the fourth column. The long-run average comparison is found by summing the intercept and beta coefficient in Table 24. For turnover ratio, the funds showed less gambling behavior than their own long-run average; they had fewer redundant trades than their lifetime annual average. For the change in portfolio concentration, right tail funds increased their portfolio concentration slightly more than their long-run average, but the difference between them and other funds is not economically or statistically significant.

Next, I investigate if the funds modified their beta exposures more than their own long-run average. The left hand side variable as absolute changes in factor exposures of each fund in the last available holdings observation before the one which generated an extreme return. In other words, the left hand side variable is the absolute value of the difference of factor exposures between two holdings observations before the return was realized and one holdings observation before the return was realized.

First, results in Table 27 document the long-run behavior of the change in absolute factor exposures in right tail funds. This provides a baseline for how much the funds on average change factor exposures between holdings over their lifetime. The measure is very similar to the standard deviation of factor exposures over funds' lifetime in Table 25, and the results of Table 27 show an identical pattern; across their lifetime, right tail funds vary their factor exposures between holdings more than the average fund. However, the previous measure, the σ of factor exposures, is more difficult to compare to fund behavior in one specific holdings observation.

The changes in absolute factor exposures just before the right tail return realization are presented in Table 28. Note that Thompson standard errors are no longer needed for this regression because the left hand side variable does not have overlapping measurement periods. Time fixed effects are also no longer necessary, and since I'm trying to compare two averages (the long run average change of factors by funds versus their average change just before realization). Similarly, the expense and turnover ratio controls are also unnecessary. For simplicity, only the beta for the most extreme group $R^{ex} \geq 15\%$ is reported alongside as the regression intercept. The sum of the intercept and beta will be compared against the long-term changes of factor exposures calculated for the right tail funds.

Table 28 displays each seven factor exposure in the first column. In the second and third columns, the intercept and beta from the Equation 18 are reported, with statistical significance recorded for β_1. The third column reports the sum of the intercept and β_1, which is the average absolute value of the change in factor exposures from the holdings observation before the right tail return was realized. The third column should be compared against the fourth column, which is the sum of the intercept and β_1 from Table 27, which reports the average absolute value of change in factor exposures for right tail funds across their lifetime. If the sum in the third column is higher than the sum in the fourth column, the funds were varying their factor exposures more than their own long-run average in the period before they

Table 27

Long-run average abs. value of change in factor exposures

	α	β_1
MKT	0.048	0.026***
SMB	0.058	0.041***
HML	0.072	0.035***
UMD	0.040	0.031***
RMW	0.88	0.055***
CMA	0.096	0.056***
STR	0.0004	0.0003***

Table 27 displays results of regressions of measures of transient strategies on a dummy variable for extreme right tail performers. The left hand side variable is the absolute change of factor exposures between holdings observations for the fund's lifetime. The first column reports the regression intercept and the second reports the beta on the right tail dummy. Statistical significance is shown at the usual thresholds: 10%, 5%, and 1% for one, two and three stars, respectively.

Table 28

Average abs. value of change in factor exposures before return

	α	β_1	Sum	Comparison: LR average
MKT	0.044	0.109***	0.153	0.074
SMB	0.055	0.072***	0.127	0.099
HML	0.067	0.103***	0.170	0.107
UMD	0.036	0.077***	0.113	0.071
RMW	0.081	0.094***	0.175	0.143
CMA	0.089	0.134***	0.223	0.152
STR	0.0004	0.001***	0.0014	0.0007

Table 28 displays results of regressions of change in factor exposures on a dummy variable for extreme right tail performers. The left hand side variable is the change in factor exposures in the holdings observation matched with funds that have contemporaneous excess return greater than 15%. The first column reports the regression intercept, the second reports the beta on the right tail dummy, and the third reports the sum of the coefficients (which is the average of these variables in right tail funds). The fourth column reports the long run average of these variables for all fund. Statistical significance is shown at the usual thresholds: 10%, 5%, and 1% for one, two and three stars, respectively.

earned their excessive return.

The table presents a uniform pattern: the sum in the third column, representing short-term behavior, is greater than their own long run average change in factors reported in the fourth column. Funds vary their factor exposures a lot more than even their own long run average just before they realize their right tail return realization. The magnitude depends on which factor is being considered but the variation is typically between 50% and 100% more than their own long run averages, which are in turn 50% to 100% greater than the long run averages of the average fund. Right tail funds have very inconsistent factor exposures and appear to gamble with them via large changes in exposures just before they receive their spectacularly high, and misleading, positive return.

1.4.4 Identifying gamblers in the full sample

The results of this study present evidence that funds in the extreme right tail of excess returns are poor quality managers. However, the result applies to only a few funds. 174 fund-date observations across 20 years are identified as gamblers. The average size of these funds is high, meaning they attract significant positive inflows in the month following their realization. A quick 'back of the envelope' calculation shows that 174 funds with an average size of 518.66m and 3.623% inflows in the month following their realization have attracted \$3.266 billion of inflows over 20 years of the sample period, all of which has reversed after 12 months.

However, the right tail only contains gamblers that were successful with their gambling behavior. Not all funds that gamble are lucky enough to get an extreme right tail realization. As a first pass at identifying other gambling funds I set a binary variable equal to 1 if a fund has (1) at least as high average absolute value of factor exposures across all seven factors, (2) at least as high standard deviation of factor exposures across all seven factors, and (3) at least as high turnover ratio as the typical extreme right tail funds across the fund's lifespan. In other words, I identify funds with (i) as extreme average strategies and (ii) as extreme variance of strategies as the typical extreme right tail fund. Other funds have a value of 0 for this dummy. Additionally, I set this variable equal to 0 for funds in the extreme right tail to ensure results are not being driven by funds already documented as poor quality gamblers. 656 additional gambling funds are identified using this methodology - as a reminder, the full sample has 5,818 unique funds, 4,435 of which have at least one high quality holdings observation. Concretely, for different measures of performance $y_{i,t}$, I run the following regression:

$$y_{i,t} = \alpha + \beta_1 \mathbb{1}_{Gamble} + Controls_{i,t} + timeFE_t + \epsilon_{i,t} \tag{19}$$

The regression is run twice with the y variable as cumulative excess return and average monthly seven factor alpha over the following year. Table 29 and 30 present results of β_1 for 12 month cumulative excess return and 12 month average alpha, respectively. As always, the first column reports the full sample, while the second includes turnover and expense ratios as controls.

Although the sorting method is preliminary, the results look promising. When 12 month cumulative excess return is used as the left hand side variable in Table 29, funds classified as gamblers underperform other funds by 3.680% and 2.856% cumulative excess return in the following year respectively for the two control variable specifications. Both results are statistically significant at the 1% level, although their underperformance is less severe than the estimate of cumulative excess return for funds sorted into the extreme right tail. When monthly seven factor alpha is considered as the left hand side variable in Table 30, the underperformance of gambling funds is statistically significant, and is economically as significant as the results

Table 29
Cumulative excess return over the following year by gamble dummy

	(1)	(2)
Gamble	-3.680***	-2.856***
Age	0.63***	0.458***
Size	-0.041*	0.016
Expense Ratio		0.779**
Turnover Ratio		-0.515***
N	586,459	468,169
Time FE	yes	yes
Thompson standard errors	yes	yes

Table 29 displays the beta of the 'gamble' dummy variable with a regression of cumulative excess return in the 12 months following bucket formation on return buckets and controls. Stars indicate statistical significance at 10%, 5%, and 1% thresholds for one, two and three stars, respectively. While both columns have age and size control variables, the second column adds expense and turnover ratio variables as controls.

Table 30
Average monthly 7F alpha over the following year by gamble dummy

	(1)	(2)
Gamble	-0.231***	-0.539***
Age	0.034***	0.033***
Size	-0.005	-0.007**
Expense Ratio		0.007
Turnover Ratio		-0.035
N	256,868	230,763
Time FE	yes	yes
Thompson standard errors	yes	yes

Table 30 displays the beta of the gamble dummy variable from a regression of 7 factor average monthly alpha for the 12 months following bucket formation on return buckets and controls. Stars indicate statistical significance at 10%, 5%, and 1% thresholds for one, two and three stars, respectively. While both columns have age and size control variables, the second column adds expense and turnover ratio variables as controls.

found from the extreme right tail funds. In the specification without expense and turnover ratios as controls, funds generate a statistically significant alpha of -24 basis points compared to non-gambling funds. The average monthly alpha decreases to -54 basis points when expense and turnover ratios are included as controls, a surprising reversal of previously reported results where expense and turnover ratio tend to positively bias the sample. For comparison, extreme right tail funds reported earlier in this paper (in Table 17) underperformed the omitted group by 24 and 23 basis points for the vector of controls with just size and age and the vector of controls with size, age, and turnover and expense variables, respectively.

1.5 Robustness

In the robustness section I examine several variations of the hypothesis and the sample size. First, I discuss how results change when the dot com bubble is included. Second, I consider Morningstar ratings being included as controls. Third, I consider fund performance over other horizons than one year. Fourth, I present results from a Locally Weighted Regression to attempt to discern a more precise threshold above which positive return is a poor predictor of future performance. Finally, I discuss the implementation of a trading strategy based on right tail funds.

1.5.1 The Dot Com Bubble

The original date range for this study was 1990-2023, which encompasses several dramatic events for the US economy. The most severe of these for my purposes was the dot com bubble.

The dot com bubble began building around 1995 and culminated in March 2000. The post-bubble trough was in October 2002, when the NASDAQ had lost all of its gains from 1995-2000. During the dot com bubble, the NASDAQ first spectacularly outperformed the market, then spectacularly underperformed the market. During this time many funds were experiencing tail returns simply by holding any amount of tech stocks.

This is problematic for my question and for my analysis. Due to the overwhelming amount of outperforming and underperforming funds during this period, by including a date range which includes the dot com bubble my investigation becomes a study about the dot com bubble, not about mutual fund behavior more generally. For example, of the 1,156 fund-month observations with excess return greater than 15% from 1990-2023, 982 of them come from the dot com bubble period of 1995-2003, versus 174 fund-month observations in other periods.

Even if receiving high excess returns during the dot com bubble was the same signal of managerial quality as in more normal times including such a large number of tech funds in the right tail makes the external validity of results questionable. However, results on performance do vary when the dot com bubble is included in the sample period.

Tables in the appendix detail how dot com bubble results differ from full sample results. Table 34 reports average factor exposures of right tail funds from 1990-2023, Table 35 reports total counts and unique dates of funds in each return bucket, Figure 9 reports the count of observations in the extreme right tail at different dates, Table 36 reports results on cumulative excess returns, Table 37 reports results on Sharpe Ratio, and Tables 38 and 39 present results on three and seven factor alphas for years 1990-2003. Similar to the main study, performance for each measure is evaluated one year after bucket formation. To summarize the results of the tables, including the dot com bubble overwhelmingly weighs the right tail sample towards funds in holding small growth stocks. Funds sorting into either tail is a monotonically bad signal of future performance in terms of excess return and the Sharpe Ratio of funds; the further towards either tail a fund is sorted into, the worse their performance over the next year. The magnitudes are even more extreme than what were reported in the post 2003 results previously detailed in Section 3, as the NASDAQ had dramatically low returns after the bubble popped. The largest difference in the full sample results versus post 2003 results is in alpha. When the dot com bubble is included in the sample, funds in the right tail have positive monthly alphas even as they are dramatically underperforming the market.

Similar to the analysis of left tail performance earlier in the paper, there are a few possible interpretations of the positive alpha result. First, results in this framework are an upper bound of true performance due to selection bias in requiring high quality holdings information. Second, and uniquely to the dot com bubble, it may be that my measurement of factor exposures - and particularly, of cash holdings - are not frequent enough. My variable of cash holdings, which comes from CRSP, is updated at an annual frequency. Any fund that was holding more cash than their yearly average when the bubble popped would have outperformed their exposures by the mere fact they have some unobserved portion of their portfolio that is zero beta when the factor realizations of SMB and HML are extremely bad for their holdings. This can optimistically be interpreted as skill, but it is also possible that managers were simply unable to quickly spend their enormous inflows of cash they received while they bubble was building.

Other papers, for example, Kosowski (2011), focus explicitly on mutual fund manager performance during recessions vs. expansions, which may be applicable to the interpretation of the positive alpha of right tail funds in the dot com bubble. However, further analysis of this single dramatic event is outside the scope of the initial hypothesis of this paper.

1.5.2 Using Morningstar rating as a control variable

In the Data section I argued that using Morningstar ratings as a control variable in regressions biased results too high to consider using them in my study. To investigate that claim, ratings are matched by WFICN between CRSP and the Morningstar database and aggregated to fund level by taking the weighted average of share class level Morningstar ratings. Some funds have non-integer Morningstar ratings when a more expensive share class has a lower rating compared to a cheaper share class within the same fund.

I will present the results of this bias in several ways. First, Table 31 will do the same regression measuring biases for the expense and turnover ratios as well as requiring high quality holdings information. I simply measure how much the dummy for ratings information is related to the fund's size and one-year excess return. Second, I will investigate excess return results across four specifications: (1) with just size and age controls, (2) adding expense and turnover ratios (i.e. what was previously presented in the results section), (3) with Morningstar ratings as a control variable, and (4) without Morningstar rating as an explicit control variable - that is, the second specification - but first dropping observations that don't have a Morningstar rating.

First, I present how the presence of a Morningstar rating is correlated to size and next year's performance. 469,013 fund-date observations have ratings information, while 190,721 fund-date observations do not.

Table 31

Selection introduced by requiring Morningstar ratings

	Full sample	$R^{ex} \leq -15\%$	$R_{ex} > 15\%$
TNA	1202.57***	1177.733***	82.896
$R^{ex}_{t\in 1,12}$	1.084***	10.531***	20.496***
Time FE	yes	yes	yes

Table 31 compares the total net assets and cumulative twelve month excess return between subsets of observations. In the first column, the average of TNA and return are computed for funds which have Morningstar ratings. This should be compared with the full sample averages reported in Table 1. In the second column, the average of these variables for extreme left tail performers with a non-missing Morningstar rating is shown. In the third column, the average of these variables for extreme right tail performers with a non-missing Morningstar rating is shown. Stars indicate statistical significance at 10%, 5%, and 1% thresholds for one, two and three stars, respectively.

Morningstar ratings introduce the largest bias in terms of size, with funds that include ratings information being 1202.57m larger than the average fund without a rating in the full sample. The upward bias of size is about the same magnitude when left tail funds are considered but much smaller for right tail funds: the average right tail fund with Morningstar ratings is only about 83 million larger than other right tail funds. The bias for next year's return in the full sample is slightly higher than the bias introduced by including expense and turnover ratios but below the bias found when requiring funds to have a high quality holdings match. However, although the bias is not too strong for expected returns in the full sample, the bias increases dramatically across the most extreme groups in either tail. Funds in the extreme left and right tail with ratings information have a staggering average of 10.531% and 20.496% higher cumulative excess return in the following year compared to their peers in the tail without ratings information.

To show how this interferes with performance results presented earlier in the paper, Table 32 presents results with 12 month excess return as the dependent variable and two additional columns showing results when only funds with ratings are selected.

Table 32 shows the extreme positive selection in the extreme tails induced by including Morningstar ratings as controls. The results in columns (1) and (2) are

Table 32
Cumulative excess return over the following year by return bucket

	(1)	(2)	(3)	(4)
$R^{ex} < -15\%$	-0.228	2.870	11.881	12.812
$-15\% \leq R^{ex} < -10\%$	-2.15	-2.273	0.767	0.700
$-10\% \leq R^{ex} < -5\%$	-2.282**	-2.071**	-1.662	-1.771*
$-5\% \leq R^{ex} < -2\%$	-0.291	-0.235	-0.138	-0.125
$2\% \leq R^{ex} < 5\%$	0.696	0.720	0.726**	0.663*
$5\% \leq R^{ex} < 10\%$	0.735	0.805	0.797	0.617
$10\% \leq R^{ex} < 15\%$	0.791	0.022	2.406	2.166
$R^{ex} \geq 15\%$	-10.134**	-8.202	1.774	1.928
Age	0.6727***	0.397**	0.588**	0.685***
Size	-0.047	0.031	0.027	-0.005
Expense Ratio		0.752**	0.900**	0.900**
Turnover Ratio		-0.516***	-0.523***	-0.535***
Rating			-0.081	
N	621,268	495,148	412,692	412,692
Time FE	yes	yes	yes	yes
Thompson standard errors	yes	yes	yes	yes

Table 32 displays the beta of each return bucket from a regression of cumulative excess return in the 12 months following bucket formation on return buckets and controls. Stars indicate statistical significance at 10%, 5%, and 1% thresholds for one, two and three stars, respectively. The first column displays the results with just age and size variables as controls, while the second adds expense and turnover ratio controls. The third column additionally adds Morningstar rating as a control. The fourth column uses the specification of the second column - age, size, expense and turnover ratio controls - but on the subset of funds that have non-missing Morningstar ratings.

identical to Table 14, which shows the performance of buckets over the following year as measured by cumulative excess return. Column (3) adds Morningstar ratings to the vector of controls. Despite the estimate of rating not being statistically significant, the results in the left and right tails become extremely positively biased; the beta estimate of the left tail has increased from 2.870% to 11.881%, and the extreme right tail has increased from -8.202% to 1.774%.

Column (4) confirms that the tail results in column (3) are driven by selection instead of the inclusion of the control itself. Column (4) reports the results of specification (2) but using on the sample of funds that have non-missing ratings information. The results show an almost identical upward bias, confirming the selection of tail funds is what is driving their sudden positive performance. Mornigstar ratings are too positively selected in the tails to consider using as a control variable.

1.5.3 Other horizons

The Appendix presents tables on fund performance by bucket over six month, two year, three year, and five year horizons. Table 40 presents excess return over other horizons, Table 41 presents Sharpe Ratio over other horizons, and Tables 42 and 43 results on three and seven factor alphas over longer horizons. The results are almost entirely identical with results presented earlier in the paper. The only specification where the extreme right tail has a positive association with performance is over six month horizons with monthly three and seven factor alphas, where betas are positive but not statistically significant. In two, three, and five year horizons performance of the extreme right tail is always negative, although sometimes without significance. By requiring longer survival of funds after sorting the count of funds in the extreme right tail diminishes over time, which increases the standard deviation of estimated betas.

The improvement of performance in the extreme left tail varies considerably between specifications. However, similar to the one year horizon results, they nearly always outperform the -15% to -10% group, with the exception of specifications where Sharpe Ratio is used as the dependent variable.

1.5.4 Locally Weighted Regressions

The formation of buckets is in a sense both arbitrary and overly rigid at the margins. A fund beating the market by 5.01% to shouldn't perform very differently from a fund beating the market by 4.99%, and so on. This section aims to investigate the association between current excess return and future fund performance in a more continuous fashion, utilizing Locally Weighted Regressions (LWRs) (Cleveland and Devlin (1988)). The fitting technique requires running many overlapping regressions in different ranges of the x variable, in this case, the fund's contemporaneous excess return.

In order to include controls in the estimation, the plotted y variable (cumulative 12 month excess return from t+1 to t+12) is first regressed on controls - age, size, expense and turnover ratios, and date fixed effects. The LWR fit is then run with residuals from the first stage regressions as the left hand side variable and contemporaneous excess return as the right hand side variable. Results of the LWR smoother is shown in Figure 8.

Figure 8 shows the LWR fit of contemporaneous excess return against cumulative excess return in the following year. In order to more closely examine the fit line the y axis has limits between -10% and 2% cumulative excess return in the following year: many individual funds have cumulative excess returns outside of this range and are not shown, but a wider y axis range visually obscures detail in the LWR fit. Performance of return in either tail is associated with monotonically worse cumulative excess return in the following 12 months. The fit is slightly positive in a narrow region between 0 and 3.56% contemporaneous excess return, where funds beat the market in the coming year. After the fit line crosses the y axis at positive 3.56% the fit is monotonically negative into the extreme right tail. The fit supports the initial hypothesis that bad managers can be found in both the left and right

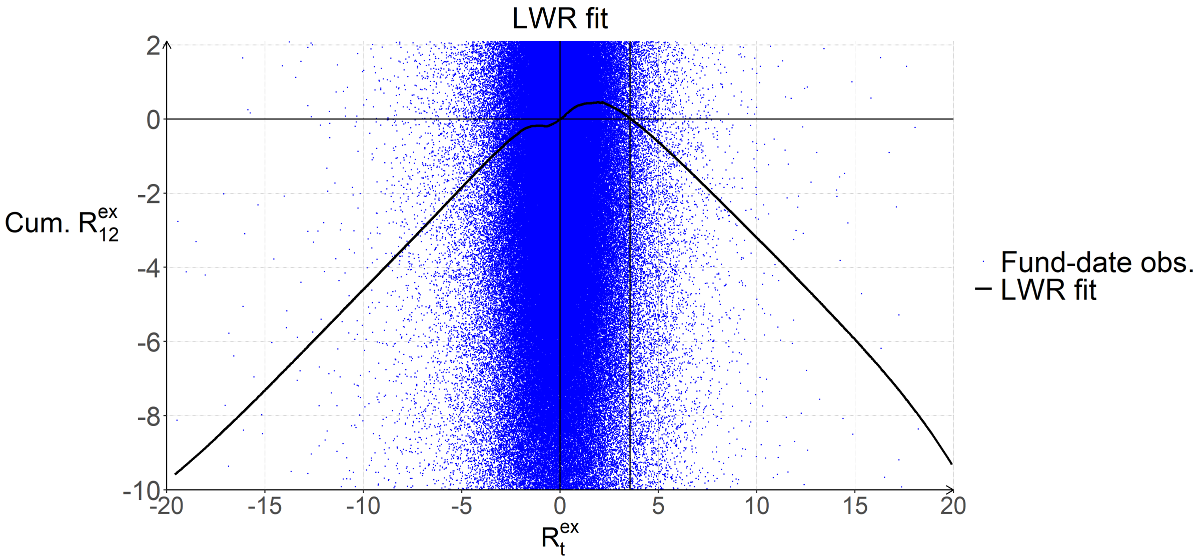

Figure 8. The x axis shows a fund's contemporaneous excess return, while the y axis shows the fund's cumulative excess return over the following year (t+1 to t+12). Blue dots correspond to an individual fund-month observation, while the black line displays the locally weighted regression fit with a bandwidth of 0.6. The vertical line identifying where the estimated fit crosses the x axis is at 3.56% excess return. Cumulative excess return over the following 12 months has been adjusted by control variables in a prior regression.

tails of the return distribution.

The main downside of the LWR analysis in its current form is that I have not been able to calculate confidence intervals. In future versions of the paper I plan to bootstrap the LWR process to construct confidence intervals so I can discern if the positive performance between 0% and 3.56% is statistically significant, and find an exact threshold where more positive contemporaneous return will predict negative future performance with statistical significance.

1.5.5 Trading strategy based on right tail funds

In this section I augment the results by providing a negatively performing trading strategy based on holding funds with extreme right tail returns. A trading strategy does not feature largely in my results because there are relatively few months with extreme right tail performing funds: only 60 months out of 251 months in the sample period have any funds beating the market by over 15%. A trading strategy does not add much evidence to the hypothesis that funds with extreme right tail funds are unskilled. However, it would be reassuring if the trading strategy generated a negative alpha.

For every month t in the sample, the strategy buys all funds that received an excess return greater than 15% in the prior month $t-1$. One strategy examines buying funds at equal weights and the second value weights purchases by their total net assets in order to ensure results are not being driven by small funds. Each month's portfolio is held for 12 months. This occurs in overlapping intervals - for example, t to $t+12$, $t+1$ to $t+13$, and so on until $t+11$ to $t+23$. At $t+12$

the first portfolio has been held for a year and is sold, with proceeds buying a new portfolio of extreme right tail funds from $t+11$. Each overlapping month's portfolio receives 1/12th of the trading strategy's overall weight. Measurement of the trading strategy does not begin until 12 overlapping portfolios are able to be formed: that is, I do not record trading strategy returns in the first 12 months of the sample. If no fund exceeded the market by more than 15% within any given month the trading strategy buys the risk free asset r_f to avoid interfering with the interpretation of alpha.[8]

Table 33 summarizes results of the trading strategies for the sample period. Alpha is measured using regressions of fund returns on factors in a 12 month rolling time window, since the strategy completely changes its funds every 12 months.

Table 33

Trading Strategy Performance

	Equal Weighted	Value Weighted
Mean monthly α_{7F}	-0.16	-0.15

Table 33 reports results of a trading strategy formed from holding portfolios of funds in the extreme right tail. The equal weighted strategy holds month portfolios of extreme right tail funds with equal weight, while the value weighted strategy weights funds by their TNA. With both methodologies each month portfolio is held for 12 months before being sold. The trading strategy has 12 concurrent month portfolios at any point in time, each receiving 1/12th of the strategy's weight.

The results in Table 33 show that the strategy does generate a negative seven factor monthly alpha. The equal and value weighted strategies deliver very similar alphas of around -15 basis points per month. This is slightly below the average monthly seven factor alpha of extreme right tail funds in the 12 months following bucket formation shown earlier in the paper in Table 17, where I reported a -23 basis point average monthly alpha. Due to the small number of dates with extreme right tail funds I do not view this as strong evidence in favor of the hypothesis that managers in the extreme right tail of the short-term return distribution are poorly skilled. Nonetheless, it is reassuring that the strategy yields a negative monthly alpha.

1.6 Conclusion

When mutual funds are sorted on absolute instead of relative thresholds for excess return, I find evidence that right tail funds underperform. Surprisingly, they do so

[8] I also ran the trading strategy where no new funds were bought if there were no extreme right tail funds in the previous month, with all active month portfolios re-weighted equally. For example, if there were 2 months in the past 12 months where funds beat the market by more than 15%, each month's portfolio would have a weight of 1/2. There is no 12 month span in the data with zero funds in the extreme right tail, so this implementation of the strategy always holds at least one portfolio of extreme right tail funds. Results were nearly identical to the strategy here, with the equal weighted strategy earning a -15bp monthly alpha and the value weighted earning a -14bp monthly alpha over the sample.

more than left tail funds, which exhibit evidence of changing strategies to improve performance. Funds beating the market by more than 15% in the current month have poor future performance across a variety of measures: they have extreme negative cumulative excess returns, negative Sharpe Ratios, and using a novel technique that identifies factor exposures of funds from their holdings information, negative three and seven factor alphas. Finding factor exposures from holdings has a variety of benefits, allowing for tracking of fund strategy with higher frequency compared to the conventional method of regressing fund returns on factors over one to five year horizons. Additionally, it allows an insight of fund strategy before they have one to five years of consecutive returns information required by the conventional method of estimating betas by regressing fund returns on factors. The novel method of measuring factor exposures allows a higher sample size of fund alphas to study and allows insight into how quickly funds vary their strategies over time, which has implications beyond my initial hypothesis.

Funds in the extreme right tail pursue extreme strategies measured by the absolute value of their factor exposures. Crucially, these funds also *change* strategies at much higher frequencies than other funds. They gamble by trading more frequently than other funds in stocks that differ from the average factor exposures of their portfolio. Although relatively few unique funds (107) are identified as gamblers in the extreme right tail, by applying the thresholds of extreme and transient strategies originally identified from the right tail, 656 additional funds are found to be 'gamblers' in the sample. These funds also have poor performance as measured in excess return and alpha in the following year.

Although the focus of this paper has been on mutual funds, the intuition of the hypothesis can easily be extended to other agents. Any agent that beats their benchmarks by suspiciously high margins reveals themselves to have mechanically taken extremely high levels of risk. If they continue to act in a gambling fashion they will almost certainly lose everything they gained by their lucky realization.

1.7 Appendix: Betting with Beta

1.7.1 Tables: 1990-2023 sample (including the dot com bubble)

First, I run the following regression with factor exposures as the dependent variable to see if a pattern emerges in the factor exposures of funds. The cutoff for the dummy variable in this first table is the same as in the post-2003 results, with funds earning an excess return greater than 15% being classified as right tail funds. As a reminder, each fund has one observation for this specification - its within-fund average of each exposure.

$$y_i = \alpha + \beta_1 \mathbb{1}^{RT} + \epsilon_{i,t} \tag{20}$$

Table 34
Average seven factor exposures in the right tail

	α	β_1
MKT	1.00	0.103***
SMB	0.203	0.369***
HML	0.036	-0.147***
UMD	-0.051	-0.016***
RMW	0.052	-0.115***
CMA	-0.079	-0.046***
STR	0.0003	0.0000

Table 34 displays results of regressions of average value of seven factor exposures on a dummy variable for extreme right tail performers. The first column reports the regression intercept, while the second reports the beta on the right tail dummy.

Compared to the average exposures of the post-2003 sample factor exposures, the funds here have a much clearer overall pattern of style. Excluding MKT and STR, the post-2003 right tail funds only had a statistically significant pattern in 3/5 factors. In the sample with the dot com bubble, patterns emerge in all five of the factors. These are overwhelmingly weighted towards tech stocks: funds are invested in small, growth, weak profitability, and aggressively investing firms.

Table 35 summarizes the characteristics of the return buckets in the full sample. Note that a new category has been added in the full sample: there are so many funds dramatically out or underperforming the market there are enough observations to run performance analysis on a -20% to -15% and 15% to 20% bucket of excess returns. The extreme left and right tail categories have been shifted to less than -20% and greater than 20%. There are 407 unique dates total. The -5% to -2% and 2% to 5% categories do not have representation in the full 407 unique dates due to the sparsity of monthly fund data from the early 1990s. Most funds appear to have quarterly, not monthly, return data in CRSP for the first few years of the 1990s.

The Figure 9 documents the count of funds in the extreme right tail at each date. Here, the overwhelming presence of the dot com bubble in the sample is visually

Table 35
Return buckets

R^{ex}	Fund-month count	Unique dates represented
$R^{ex} < -20\%$	234	45
$-20\% \leq R^{ex} < -15\%$	544	97
$-15\% \leq R^{ex} < -10\%$	2,487	202
$-10\% \leq R^{ex} < -5\%$	22,348	379
$-5\% \leq R^{ex} < -2\%$	104,792	405
$-2\% \leq R^{ex} < 2\%$	599,790	407
$2\% \leq R^{ex} < 5\%$	95,688	404
$5\% \leq R^{ex} < 10\%$	20,969	373
$10\% \leq R^{ex} < 15\%$	2,309	180
$15\% \leq R^{ex} < 20\%$	573	92
$R^{ex} \geq 20\%$	583	49

Table 35 displays the number of fund-month observations and unique dates represented within each bucket created by sorting funds on their contemporaneous pre-fee excess return for the full sample of 1990-2023.

obvious.

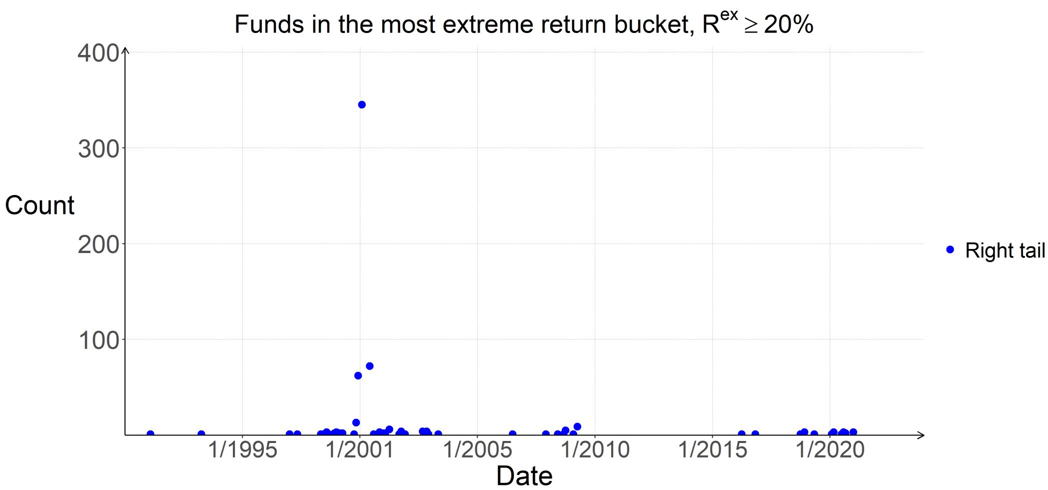

Figure 9. This figure shows the number of funds that appear in the most extreme right tail bucket in every month, $R^{ex} \geq 20\%$. The overwhelming weight of the sample towards the dot com bubble is evident, with over half of total right tail realizations occurring in February 2001.

The following tables document the future one year performance of funds in different return buckets from 1990-2023. The regression equation is

$$y_{i,t} = \alpha + \sum_{i=1}^{11} \beta_i \mathbb{1}_{(a_j < \sigma < b_j)} + Controls_{i,t} + timeFE_t + \epsilon_{i,t} \tag{21}$$

As before, the first column reports the full sample results of every fund-month observation with a return value. The second column includes turnover and expense ratios as controls.

Table 36
Cumulative excess return over the following year by return bucket

	(1)	(2)
$R^{ex} < -20\%$	-29.593***	-31.084***
$-20\% \leq R^{ex} < -15\%$	-14.896***	-15.513*
$-15\% \leq R^{ex} < -10\%$	-8.056*	-8.475*
$-10\% \leq R^{ex} < -5\%$	-2.410*	-2.093
$-5\% \leq R^{ex} < -2\%$	-0.775*	-0.719
$2\% \leq R^{ex} < 5\%$	1.482**	1.712**
$5\% \leq R^{ex} < 10\%$	2.850*	3.503**
$10\% \leq R^{ex} < 15\%$	0.303	0.240
$15\% \leq R^{ex} < 20\%$	-11.076**	-10.634**
$R^{ex} \geq 20\%$	-29.653**	-29.617***
Age	0.494**	0.307*
Size	-0.180	-0.131
Expense Ratio		0.785**
Turnover Ratio		-0.294
N	765,611	606,023
Time FE	yes	yes
Thompson standard errors	yes	yes

Table 36 displays the beta of each return bucket from a regression of cumulative excess return in the 12 months following bucket formation on return buckets and controls in the full sample from 1990-2023. Stars indicate statistical significance at 10%, 5%, and 1% thresholds for one, two and three stars, respectively. While both columns have age and size control variables, the second column adds expense and turnover ratio variables as controls.

Table 37

Annualized Sharpe Ratio over the following year by return bucket

	(1)	(2)
$R^{ex} < -20\%$	-0.694***	-0.571***
$-20\% \leq R^{ex} < -15\%$	-0.490***	-0.382***
$-15\% \leq R^{ex} < -10\%$	-0.389***	-0.307***
$-10\% \leq R^{ex} < -5\%$	-0.200***	-0.144**
$-5\% \leq R^{ex} < -2\%$	-0.104***	-0.090***
$2\% \leq R^{ex} < 5\%$	-0.207	0.003
$5\% \leq R^{ex} < 10\%$	-0.033	0.022
$10\% \leq R^{ex} < 15\%$	-0.218*	-0.122
$15\% \leq R^{ex} < 20\%$	-0.517***	-0.413***
$R^{ex} \geq 20\%$	-0.867***	-0.776***
Age	0.002	0.003
Size	0.004	-0.003
Expense Ratio		-0.046*
Turnover Ratio		-0.033**
N	761,948	603,423
Time FE	yes	yes
Thompson standard errors	yes	yes

Table 37 displays the beta of each return bucket from a regression of Sharpe Ratio in the 12 months following bucket formation on return buckets and controls for the full sample from 1990-2023. Stars indicate statistical significance at 10%, 5%, and 1% thresholds for one, two and three stars, respectively. While both columns have age and size control variables, the second column adds expense and turnover ratio variables as controls.

Table 38
Average monthly 3F alpha over the following year by return bucket

	(1)	(2)
$R^{ex} < -20\%$	-0.018	0.094
$-20\% \leq R^{ex} < -15\%$	-0.170	-0.105
$-15\% \leq R^{ex} < -10\%$	-0.492***	-0.443***
$-10\% \leq R^{ex} < -5\%$	-0.228***	-0.223***
$-5\% \leq R^{ex} < -2\%$	-0.088***	-0.085***
$2\% \leq R^{ex} < 5\%$	0.016	0.025
$5\% \leq R^{ex} < 10\%$	-0.007	0.013
$10\% \leq R^{ex} < 15\%$	0.122	0.186
$15\% \leq R^{ex} < 20\%$	0.341	0.399
$R^{ex} \geq 20\%$	0.796***	0.871***
Age	0.004	0.010
Size	-0.009***	-0.014***
Expense Ratio		-0.018
Turnover Ratio		0.002
N	360,775	311,212
Time FE	yes	yes
Thompson standard errors	yes	yes

Table 38 displays the beta of each return bucket from a regression of 3 factor average monthly alpha for the 12 months following bucket formation on return buckets and controls for the full sample from 1990-2023. Stars indicate statistical significance at 10%, 5%, and 1% thresholds for one, two and three stars, respectively. While both columns have age and size control variables, the second column adds expense and turnover ratio variables as controls.

Table 39
Average monthly 7F alpha over the following year by return bucket

	(1)	(2)
$R^{ex} < -20\%$	0.605**	0.781***
$-20\% \leq R^{ex} < -15\%$	0.203	0.324***
$-15\% \leq R^{ex} < -10\%$	-0.233***	-0.179***
$-10\% \leq R^{ex} < -5\%$	-0.082***	-0.085***
$-5\% \leq R^{ex} < -2\%$	-0.060***	-0.064***
$2\% \leq R^{ex} < 5\%$	0.067*	0.068**
$5\% \leq R^{ex} < 10\%$	0.136	0.143
$10\% \leq R^{ex} < 15\%$	0.379	0.413*
$15\% \leq R^{ex} < 20\%$	0.799**	0.810***
$R^{ex} \geq 20\%$	1.479***	1.502***
Age	0.007	0.015
Size	-0.008**	-0.010**
Expense Ratio		0.01
Turnover Ratio		-0.008
N	360,775	311,212
Time FE	yes	yes
Thompson standard errors	yes	yes

Table 39 displays the beta of each return bucket from a regression of 7 factor average monthly alpha for the 12 months following bucket formation on return buckets and controls for the full sample from 1990-2023. Stars indicate statistical significance at 10%, 5%, and 1% thresholds for one, two and three stars, respectively. While both columns have age and size control variables, the second column adds expense and turnover ratio variables as controls.

The most noticeable difference with the main results are (1) the symmetry of poor performance in the excess return measure and (2) the positive and significant alphas in the right tail. Despite underperforming the market by -30% over the year following bucket formation the extreme right tail also earns a baffling +1.5% monthly alpha, as reported in Table 39. Discussion of this result takes place in the robustness section. The closest full sample result to the post-2003 results is the regression with Sharpe Ratio as a left hand side variable. In any date range or control specification Sharpe Ratio monotonically declines the further towards either tail the funds are sorted, making it the most consistent indicator of negative performance in the right tail.

1.7.2 Tables: Performance over other horizons

Beginning in Table 40 I present results on mutual fund performance of six month, two year, three year, and five year horizons to complement the one year horizons previously recorded in the paper. The first column of each horizon reports the specification in the full sample, while the second will report the specification with expense and turnover ratios as controls. For example, 6m reports full sample results for six month performance with just age and size controls, while 6m' reports results with the expense and turnover ratios included as controls.

The results at different horizons are broadly in line with the results of the paper. The only specifications the extreme right tail has a positive beta estimate for performance is on six month horizons with monthly 3 and 7 factor alphas, but the estimates are not statistically significant. The left tail performance varies considerably but with the exception of regressions with Sharpe Ratio as the dependent variable always outperform the -15% to -10% bucket.

Table 40

Cumulative excess return by return bucket over 6m, 2y, 3y, and 5y horizons

	6m	6m'	2y	2y'	3y	3y'	5y	5y'
$R^{ex} < -15\%$	-8.955**	-7.774**	-0.257	5.053	0.268	7.040	-33.653***	-17.086**
$-15\% \leq R^{ex} < -10\%$	-4.534***	-4.919***	-2.442	-1.897	-1.026	-0.508	-11.192*	-7.942
$-10\% \leq R^{ex} < -5\%$	-1.463**	-1.395**	-1.166	-0.910	-0.724	-0.694	0.440	0.212
$-5\% \leq R^{ex} < -2\%$	-0.082	-0.058	-0.112	0.025	-0.516	-0.417	-0.299	-0.315
$2\% \leq R^{ex} < 5\%$	0.381	0.370	0.995*	0.963	1.118	1.027	1.858**	1.446*
$5\% \leq R^{ex} < 10\%$	0.681	0.728	0.307	0.392	-0.667	-0.663	0.138	-0.540
$10\% \leq R^{ex} < 15\%$	0.754	0.848	-3.204**	-2.755*	-6.244**	-5.798**	-8.819	-8.584
$R^{ex} \geq 15\%$	-2.319	-1.739	-19.761***	-16.890**	-25.147***	-20.959***	-30.750***	-22.890***
Age	0.339***	0.193**	1.303***	0.803**	1.832***	1.125**	2.752**	1.668*
Size	-0.017	0.013	-0.135	0.021	-0.205	0.042	-0.276	0.183
Expense Ratio		0.397**		1.595***		2.586***		4.245***
Turnover Ratio		-0.278***		-0.975***		-1.339***		-1.808***
N	660,262	519,167	552,300	442,243	490,124	394,027	382,276	309,871
Time FE	yes	yes	yes	yes	yes	yes	yes	yes
Thompson standard errors	yes	yes	yes	yes	yes	yes	yes	yes

Table 40 displays the beta of each return bucket from a regression of cumulative excess return over various horizons on return buckets and controls. Stars indicate statistical significance at 10%, 5%, and 1% thresholds for one, two and three stars, respectively. For each time horizon, the first column displays the beta of the full sample regression with size and age controls, while the second column adds expense and turnover ratios as control variables.

Table 41

Sharpe Ratio by return bucket over 6m, 2y, 3y, and 5y horizons

	6m	6m'	2y	2y'	3y	3y'	5y	5y'
$R^{ex} < -15\%$	-1.401***	-1.272***	-0.532	-0.319	-0.471	-0.271	-1.050***	-0.705**
$-15\% \leq R^{ex} < -10\%$	-0.849***	-0.805***	-0.346	-0.296	-0.245	-0.210	-0.517***	-0.437***
$-10\% \leq R^{ex} < -5\%$	-0.315***	-0.293***	-0.170***	-0.148***	-0.149**	-0.138**	-0.124**	-0.117**
$-5\% \leq R^{ex} < -2\%$	-0.098***	-0.036	-0.069**	-0.061**	-0.064**	-0.058**	-0.052*	-0.046
$2\% \leq R^{ex} < 5\%$	-0.047	-0.036	-0.054**	-0.047**	-0.056***	-0.049**	-0.036	-0.030
$5\% \leq R^{ex} < 10\%$	-0.161	-0.146	-0.131**	-0.125**	-0.147***	-0.137***	-0.111**	-0.010**
$10\% \leq R^{ex} < 15\%$	-0.729***	-0.632***	-0.416***	-0.366***	-0.409***	-0.366***	-0.396***	-0.357***
$R^{ex} \geq 15\%$	-0.967**	-0.751*	-0.841***	-0.710***	-0.738***	-0.590***	-0.826***	-0.621***
Age	0.0008	0.0113	0.008	0.009	0.006	0.006*	0.009***	0.009**
Size	0.011**	0.001	0.005	0.000	0.003	0.000	0.002	0.000
Expense Ratio		-0.059**		-0.038*		-0.022		-0.009
Turnover Ratio		-0.044***		-0.031***		-0.026***		-0.021***
N	602,118	474,865	519,459	416,432	458,734	369,221	353,258	287,023
Time FE	yes	yes	yes	yes	yes	yes	yes	yes
Thompson standard errors	yes	yes	yes	yes	yes	yes	yes	yes

Table 41 displays the beta of each return bucket from a regression of Sharpe Ratio over various horizons on return buckets and controls. Stars indicate statistical significance at 10%, 5%, and 1% thresholds for one, two and three stars, respectively. For each time horizon, the first column displays the beta of the full sample regression with size and age controls, while the second column adds expense and turnover ratios as control variables.

Table 42

3 factor alpha by return bucket over 6m, 2y, 3y, and 5y horizons

	6m	6m'	2y	2y'	3y	3y'	5y	5y'
$R^{ex} < -15\%$	-0.010	0.137	0.046	0.033	-0.100	-0.101	0.078***	0.101*
$-15\% \leq R^{ex} < -10\%$	-0.333***	-0.321**	-0.131	-0.121	-0.279*	-0.291*	-0.538***	-0.513***
$-10\% \leq R^{ex} < -5\%$	-0.223***	-0.203***	-0.058	-0.049	-0.094**	-0.086**	-0.131***	-0.113***
$-5\% \leq R^{ex} < -2\%$	-0.068***	-0.058**	-0.016	-0.008	-0.033	-0.025***	-0.023**	-0.017***
$2\% \leq R^{ex} < 5\%$	-0.005	0.010	0.004	0.009	-0.003	0.003	-0.005	-0.001
$5\% \leq R^{ex} < 10\%$	-0.091	-0.055	-0.046	-0.041	-0.053*	-0.048*	-0.091***	-0.079***
$10\% \leq R^{ex} < 15\%$	-0.168	-0.150	-0.188*	-0.171*	-0.125	-0.115	-0.166	-0.144
$R^{ex} \geq 15\%$	0.133	0.323	-0.418***	-0.384**	-0.421***	-0.405***	-0.608***	-0.563***
Age	0.022***	0.021***	0.022***	0.022***	0.019***	0.019***	0.012*	0.009***
Size	-0.004	-0.012***	-0.008***	-0.012***	-0.010***	-0.013***	-0.003***	-0.003***
Expense Ratio		-0.032**		-0.009		0.000		0.027
Turnover Ratio		-0.068***		-0.050*		-0.044		-0.046
N	304,488	270,429	185,136	168,259	131,461	121,540	58,265	56,329
Time FE	yes	yes	yes	yes	yes	yes	yes	yes
Thompson standard errors	yes	yes	yes	yes	yes	yes	yes	yes

Table 42 displays the beta of each return bucket from a regression of average monthly 3 factor alpha over various horizons on return buckets and controls. Stars indicate statistical significance at 10%, 5%, and 1% thresholds for one, two and three stars, respectively. For each time horizon, the first column displays the beta of the full sample regression with size and age controls, while the second column adds expense and turnover ratios as control variables.

Table 43

7 factor alpha by return bucket over 6m, 2y, 3y, and 5y horizons

	6m	6m'	2y	2y'	3y	3y'	5y	5y'
$R^{ex} < -15\%$	0.054	0.167	0.22***	0.216***	0.018	0.009	-0.104***	-0.124***
$-15\% \leq R^{ex} < -10\%$	-0.357***	-0.373***	-0.021	-0.016	-0.090	-0.105*	-0.281***	-0.293***
$-10\% \leq R^{ex} < -5\%$	-0.129**	-0.118**	0.044	0.042	0.019	0.009	-0.007	-0.024
$-5\% \leq R^{ex} < -2\%$	-0.056**	-0.051**	-0.008	-0.008	-0.010	-0.012	0.020	0.013***
$2\% \leq R^{ex} < 5\%$	0.048	0.051	0.065**	0.058*	0.047***	0.042**	0.027***	0.019***
$5\% \leq R^{ex} < 10\%$	0.005	0.021	0.110	0.091	0.074*	0.060	0.007	-0.004
$10\% \leq R^{ex} < 15\%$	0.030	0.025	-0.004	-0.035	0.030	0.006	-0.070	-0.107
$R^{ex} \geq 15\%$	0.076	0.149	-0.328	-0.346	-0.225	-0.242	-0.443***	-0.465***
Age	0.033***	0.030***	0.033***	0.033***	0.030***	0.031***	0.023***	0.022***
Size	-0.002	-0.006**	-0.010***	-0.010***	-0.012***	-0.011***	-0.005***	-0.001*
Expense Ratio		-0.006		0.016		0.022		0.047*
Turnover Ratio		-0.042*		-0.013		-0.003		0.005
N	304,488	270,429	185,136	168,259	131,461	121,540	58,265	56,329
Time FE	yes	yes	yes	yes	yes	yes	yes	yes
Thompson standard errors	yes	yes	yes	yes	yes	yes	yes	yes

Table 43 displays the beta of each return bucket from a regression of average monthly 7 factor alpha over various horizons on return buckets and controls. Stars indicate statistical significance at 10%, 5%, and 1% thresholds for one, two and three stars, respectively. For each time horizon, the first column displays the beta of the full sample regression with size and age controls, while the second column adds expense and turnover ratios as control variables.

2 Rationality of the Flow-Performance Relationship in Mutual Funds

2.1 Introduction

The evaluation of fund manager skill based on their raw return has long been a puzzle in mutual fund literature. Multiple studies have shown that investors flow into any fund with positive return and out of any fund with negative return (Chevalier and Ellison (1997), Cuthbertson et al. (2010), Ben-David et al. (2022)). Some authors have claimed that flows predict abstract and rational measures of performance such as α (Berk and Van Binsbergen (2015), Barber et al. (2016)) while others argue investors are simple agents chasing positive return without sophisticated reasoning (Ben-David et al. (2022)).

I aim to establish a rational benchmark for how flows should act when positive or negative mutual fund performance is revealed. In the model, I consider whether an investor would choose to re-hire a manager after observing their return realizations to represent how money flows into or out of mutual funds. In contrast to Berk and Green (2004), where the authors assume managers are skilled but have decreasing returns to scale, I allow for the existence of value destroying managers. The addition of an explicitly value destroying or 'bad' fund manager in my model is supported empirically (see, e.g. Cuthbertson et al. (2010)) and adds an interesting dimension to an investor's re-hiring decision: they would ideally like to only commit their money to good managers. However, bad managers can respond to any re-hiring rule which attempts to fire bad managers by attempting to imitate a good manager's exact actions. If a bad manager is lucky and the imitation succeeds, good and bad managers are indistinguishable.

To see how investors optimally re-hire managers of different quality, I investigate the welfare of an investor under different re-hiring rules. I use a baseline model from Dasgupta and Prat (2006), which is a Kyle-type model with noise traders that facilitate an informed trader making a profit with rational market makers (Kyle (1985)). In the model of Dasgupta and Prat (2006), good and bad managers make trading decisions and attempt to get re-hired by a rational investor seeking to only hire good managers. Good managers always trade correctly in the model's sole asset, since they receive a perfectly informative signal on the asset's end of period value. When career concerns are not present i.e. the investor's re-hiring rule is exogenous, the bad manager prefers to not trade - they receive no information on the end of period value of the asset, so any trade would just cost money in expectation via the market maker's spread. In contrast, when the investor designs a welfare optimizing re-hiring rule, they only re-hire a manager who trades correctly in the asset. The bad manager with career concerns then engages in costly imitation of the good manager and is re-hired on a successful guess.

Motivated by the results of my first chapter, I introduce a second asset which I call the *noise* asset alongside the original asset (which I call the *expertise* asset). Identically to the original paper, a good manager receives a perfectly precise signal on the expertise asset. In contrast, it is common knowledge that neither manager receives an informative signal about the end of period value of the noise asset. In this way, a lucky bad manager may outperform a good manager: it will be shown that the noise asset has a smaller spread than the expertise asset, so even though the two assets have an identical payoff distribution, trading correctly in the noise asset gives the manager a higher profit than trading correctly in the expertise asset. However, if a bad manager decides to trade the noise asset, the reveal their type, as a good manager always trades the expertise asset. The empirical result of the first chapter is that lucky bad managers outperform good managers and are rewarded by flows; could this be consistent with a rational investor in a theoretical model of managerial actions with career concerns?

To establish a baseline, I first investigate manager actions if the re-hiring decision of the investor is exogenous (unrelated to the manager's trading action or payoff realization). This can also be interpreted as the manager not having career concerns, i.e. the need to achieve a good performance to be re-hired. In the absence of career concerns, good managers profitably trade on their superior information in the expertise asset. Bad managers are indifferent between gambling on the noise asset, which has an expected payoff of zero, and not trading. They prefer not to trade the expertise asset because market makers set a positive bid-ask spread on it due to the possible presence of an informed trader. If a bad manager attempted to trade the same asset as good managers they would incur half the spread as a cost with certainty but would have an expected value of zero as a benefit, since they have no useful information on its end of period value.

When the model lasts for more than one time period, in the first period, a rational investor bases their re-hiring decision on the payoff of the manager. First, I consider the case where the investor cannot commit to a re-hiring rule with profitable ex-post deviations. In other words, the investor cannot credibly re-hire a bad manager that has revealed their type. Under this restriction a rational investor conditions their re-hiring on earning the *exact* payoff of a good manager who profitably trades in the expertise asset. The bad manager then has a choice of attempting to imitate the good manager in a pooling equilibrium by trading randomly in the expertise asset (which succeeds with a probability of 0.5) or continuing with their optimal behavior absent career concerns in a separating equilibrium. Since the benefits of being re-hired with probability 0.5 outweigh the cost of paying half of the spread, the bad manager engages in costly imitation of the good manager, guessing randomly on the value of the expertise asset. In the pooling equilibrium the bad manager pays half of the market maker's spread in order to attempt to imitate the good manager, which is costly for both the investor and the manager.

If the investor can commit to a re-hiring rule that has profitable ex-post deviations, they will optimally choose a simpler rule: to re-hire any manager with a positive payoff, regardless of which asset they traded. The bad manager will then trade in the noise asset instead of the expertise asset and will reveal their type at the end of the period. The investor with this rule has committed to the possibility of knowingly re-hiring a bad quality manager who has revealed their type. However, from an ex-ante perspective, both the investor and the manager are better off. A bad manager imitating the good manager pays half the bid-ask spread with certainty for any trade, even though their expected benefit is zero. The cost of the spread is split between the bad manager and investor according to their compensation contract. By allowing the bad manager to gamble in the noise asset instead of incurring a cost to imitate the good manager, both the investor and bad manager avoid splitting half of the bid-ask spread. The noise and expertise assets both have ex-random realizations, so under both rules the bad manager is fired with probability 0.5. Although the resulting rule looks simple and irrational - bad managers are re-hired even after revealing their type - the welfare of investors is higher than in the previous case where they only re-hired managers who earn the exact payoff of the good type. The simple rule of re-hiring any manager with a positive payoff is not inconsistent with rationality, even when both good and bad managers exist in an economy.

This paper is related to several strands of literature. I contribute to a large literature studying investor flows into mutual funds. Chevalier and Ellison (1997) first quantified the positive and convex flow-performance relationship. Berk and Green (2004), Berk and Van Binsbergen (2015), and Barber et al. (2016) provide arguments for how flows chasing performance are rational and discuss how they predict future fund performance (alpha). In contrast, Ben-David et al. (2022) argues that flows are based on simple rules of thumb by unsophisticated investors. I contribute to this literature by showing that the simple rule of thumb to flow into funds with positive returns is not inconsistent with rationality.

This paper also contributes to the literature studying career concerns of mutual fund managers. The model is an extension of Dasgupta and Prat (2006), which investigates how career concerns of mutual fund managers can lead bad managers to 'churn', or trade without information. In a later paper by the same authors, Dasgupta and Prat (2008) show that career concerns cause a breakdown in the information efficiency of markets. Similarly, Guerrieri and Kondor (2012) show that career concerns in mutual fund managers can magnify the reactions of prices to different shocks. Surveys on the positive and negative implications of career concerns include Bhattacharya et al. (2008) and Dasgupta et al. (2021). Although career concerns can cause problems in markets where agents condition re-hiring on credible signals of managerial quality, this paper demonstrates a situation where the negative consequences of career concerns can be mitigated by allowing for the bad manager to gamble and occasionally be re-hired.

2.2 A Model of Fund Manager Behavior

The model occurs over two discrete time periods, $t \in \{1, 2\}$. For simplicity, there is no discounting between time periods. There is a continuum of risk neutral agents (fund managers) hired by one risk neutral principal, the investor. The first fund manager at the beginning of $t = 1$ is hired randomly from the pool of fund managers and has quality $\theta \in \{g, b\}$, where g denotes a good manager and b denotes a bad manager. The ex-ante probability of a manager being good is γ, while a bad manager is hired with probability $1 - \gamma$. Investors also make the decision to re-hire the manager after payoffs from $t = 1$ are realized. If the first manager is not re-hired another manager is chosen randomly from the pool of managers and is good with probability γ. At the beginning of the paper I consider fund manager actions when the re-hiring decision is exogenous. In the later two sections I consider optimal re-hiring rules when the agent can or cannot credibly commit to a re-hiring rule that has ex-post profitable deviations.

There are two other agents in the model. A mass of $2N$ noise traders trade randomly and exogenously for liquidity reasons. The probability of each noise trader executing a trade in each period is η, meaning $2N\eta$ noise trades occur in total each period. There are $M >> N$ competitive, risk neutral market makers who facilitate trades. By having $M >> N$ I neglect the probability that any market maker receives two or more trades in a single period. Market makers set bid and ask prices for assets based on the expected value of that asset conditional on receiving a buy or sell order. Any agent wishing to trade is randomly and anonymously matched with an individual market maker.

There are two assets j in the model, ν and ξ. Each asset takes a random value $\nu, \xi \in \{0, 1\}$ at the end of each period. I assume that noise trades are split evenly between the two assets: since there are $2N\eta$ noise traders and two assets, each asset receives $N\eta$ noise trades in each period. The ex-ante probability of each asset realizing a value of 0 or 1 is 0.5, giving each asset an expected value of 0.5. Although the assets have identical value distributions and ex-ante probabilities of each value, it is common knowledge that a good manager receives a perfectly informative signal on the value of ν. Each manager receives a signal $S_i \in \{\emptyset, 0, 1\}$ about the end of period value of ν. Good managers receive $S_i = \nu$, revealing the end of period value of ν with perfect accuracy. They can use this information to profitably trade in ν with probability 1. Bad managers receive $S_i = \emptyset$, revealing no information about ν. In this way managers learn their own type by observing their signal. By not knowing their own quality until hired I rule out signalling by managers of different qualities or screening by the investor in their hiring decision. Neither manager receives any signal about the end of period value of ξ.

Each manager can make up to one trade each period between ξ and ν: the quantity restriction on trading prevents good manager, who is perfectly informed about the end of period value of ν, from submitting infinite trades. They can long (buy) the asset and purchase it from the market maker at $p^A_{j,t}$ or short (sell) it. If they short they asset they earn $p^B_{j,t}$ immediately but owe the realized value of the asset at the end of the period. There are no extra transaction fees for taking long or short positions on either asset. The manager can also choose not to trade. Payoffs are then mechanically defined as $\chi_t \in \{0, j - p^A_{j,t}, p^B_{j,t} - j\}$. The manager's payoff is zero if they do not trade and is otherwise the difference between the end of asset j's value and the bid or ask price depending on if the manager has taken a long or short position in the asset. Managers are compensated by the investor as defined by their compensation contract $\pi_t = \alpha\chi_t + \beta$. α is the variable component of compensation, while β is the fixed component of compensation. I make the following assumption about the compensation parameters:

Assumption 1 $\beta > 0$, $\alpha \to 0$

The assumption on the compensation parameters captures the empirical feature that most managers are compensated by AUM in each time period instead of explicitly on return. For a discussion of optimal endogenous contracting (i.e. optimal choice of α and β) by the investor see Dasgupta and Prat (2006).

2.2.1 Equilibrium without career concerns

In the first step of the solution I consider the investor making an exogenous re-hiring decision in $t = 1$. The exogenous re-hiring decision does not depend on the trades or payoffs of the manager in the previous period. Each manager then acts optimally without distortions to their behavior caused by career concerns. The managers' optimal actions without career concerns will also characterize their actions in the final period of the model, $t = 2$.

The first step to solving for equilibrium is to find how market makers set prices conditional on which manager is participating in each market for assets. Good and bad managers will make their decisions after observing bid-ask spreads. Finally, I will confirm that the beliefs of the market makers match the actions of the managers.

Since the market makers know that no agent has information about ξ, they will always set the prices of ξ as follows:

$$p^A_{\xi,t} = p^B_{\xi,t} = \frac{1}{2} \tag{22}$$

This holds regardless of the participation of either the good or bad manager in the market for ξ, as neither manager gets a signal about the end of period value of

ξ. The spread for this asset is $s_{\xi,t} = p^A_{\xi,t} - p^B_{\xi,t} = 0$, since market makers are risk neutral and perfectly competitive.

In contrast, if good managers trade on their signal to profit from ν, market makers need to set a positive spread in order to be compensated for the risk of trading against an informed agent. If the good manager is trading ν they will use their signal to trade ν profitably, since the variable component of their compensation $\alpha > 0$. The spread also depends on if the bad manager participates in the market for ν. Bad managers act as an additional noise trader in the market for ν, marginally reducing the spread needed by market makers to break even in expectation. The market maker then has three sets of bid and ask prices for ν depending on their belief on the actions of the good and bad manager. First, if the good manager is not trading ν, market makers set prices

$$p^A_{\nu,t} = p^B_{\nu,t} = \frac{1}{2} \tag{23}$$

This is identical to the prices market makers set for ξ under any set of manager actions. If the good agent is not trading ν every agent is a noise trader and receiving a buy or sell order reveals no information about the price. These prices hold regardless if the bad manager is participating in the market for ν or not. The spread of ν under these beliefs is identical to the spread of ξ, $s_{\nu,t} = p^A_{\nu,t} - p^B_{\nu,t} = 0$.

Second, if only the good manager is trading ν but the bad manager is (i) trading ξ or (ii) not trading, prices are

$$p^A_{\nu,t} = \frac{\gamma}{\gamma + N\eta}(1) + \frac{N\eta}{\gamma + N\eta} * \frac{1}{2} = \frac{1}{2}\frac{2\gamma + N\eta}{\gamma + N\eta} > \frac{1}{2} \tag{24}$$

$$p^B_{\nu,t} = \frac{\gamma}{\gamma + N\eta}(0) + \frac{N\eta}{\gamma + N\eta} * \frac{1}{2} = \frac{1}{2}\frac{N\eta}{\gamma + N\eta} < \frac{1}{2} \tag{25}$$

As before, perfectly competitive market makers set prices equal to the expected value of the asset conditional on receiving a buy or sell order. When market makers believe the good manager is participating in the market for ν there is probability γ that a good manager was hired from the initial pool and that their trade will reveal the final value of the asset. The spread market makers charge to break even is positive and equal to

$$s_{\nu,t} = p^A_{\nu,t} - p^B_{\nu,t} = \frac{1}{2}\frac{2\gamma + N\eta}{\gamma + N\eta} - \frac{1}{2}\frac{N\eta}{\gamma + N\eta} = \frac{\gamma}{\gamma + N\eta} \tag{26}$$

Finally, if both good and bad managers participate in the market for ν, market makers adjust their prices to account for the slightly higher probability of encountering an additional noise trade coming from the bad manager. Bid and ask prices and the spread are set as follows:

$$p^A_{\nu,t} = \frac{\gamma}{1+N\eta}(1) + \frac{1-\gamma}{1+N\eta}\frac{1}{2} + \frac{N\eta}{1+N\eta}\frac{1}{2} = \frac{1}{2}\frac{1+\gamma+N\eta}{1+N\eta} > \frac{1}{2} \tag{27}$$

$$p^B_{\nu,t} = \frac{\gamma}{1+N\eta}(0) + \frac{1-\gamma}{1+N\eta}\frac{1}{2} + \frac{N\eta}{1+N\eta}\frac{1}{2} = \frac{1}{2}\frac{1-\gamma+N\eta}{1+N\eta} < \frac{1}{2} \tag{28}$$

$$p^A_{\nu,t} - p^B_{\nu,t} = \frac{1}{2}\frac{2\gamma}{1+N\eta} = \frac{\gamma}{1+N\eta} \tag{29}$$

Relative to the beliefs where only good managers trade ν the spread has narrowed: $\frac{\gamma}{1+N\eta} < \frac{\gamma}{\gamma+N\eta}$. If the bad manager trades in the market for ν they provide a small positive externality for good managers by narrowing the spread charged for ν.

Based on these prices, Proposition 1 summarizes the unique equilibrium under these settings when managers do not have career concerns.

Proposition 1 *Without career concerns, good managers trade profitably on their signal of the end of period value of ν. Bad managers are indifferent between (i) trading randomly on ξ or (ii) not trading either asset.*

Good managers will always choose to trade profitably on their information for ν, since their compensation $\pi_t = \alpha\chi_t + \beta = \alpha(\frac{1}{2} - \frac{1}{2}s_{\nu,t}) + \beta$. This is greater than their expected payoff if they traded ξ or did not trade, $E(\pi_t) = \alpha(0) + \beta$. Since the spread is symmetrical the good manager's payoff χ_t is equivalent regardless if the manager buys the asset with final value 1 or sells the asset with final value 0.

Bad managers will be indifferent between not trading and trading ξ when the rehiring decision is exogenous and good managers profitably trade on their information in ν. In both cases, their expected compensation is $E(\pi_t) = \alpha\chi_t + \beta = \alpha(0) + \beta$. If they attempted to trade the expertise asset ν randomly, their compensation would be $E(\pi_t) = \alpha(-\frac{1}{2}(s_{\nu,t})) + \beta < \beta$. If a bad manager guessed on the value of ν they pay half the spread with certainty but only guess correctly with probability 0.5, meaning their expected benefit of trading ν is 0. If bad managers trade ξ they have zero costs but also zero expected benefits, as they earn a payoff of $\{-\frac{1}{2}, \frac{1}{2}\}$, each with probability 0.5.

Market makers set prices and spreads corresponding to the manager participation as in Equations 24, 25, and 26. After observing the spread neither manager has an incentive to change their behavior: the spread is positive, so the bad manager wants to avoid the market for ν, and the spread is below 1, meaning the good manager will make a profit from using their signal to trade ν. The good manager earns payoff $\chi_t = \frac{1}{2}\frac{N\eta}{\gamma+N\eta}$ with certainty and is compensated $\pi_t = \alpha\chi_t + \beta = \alpha(\frac{1}{2}\frac{N\eta}{\gamma+N\eta}) + \beta$. Bad managers earn payoff $\chi_t \in \{-\frac{1}{2}, \frac{1}{2}\}$ if they trade ξ, or $\chi_t = 0$ if they do not trade. They earn zero expected payoff with either action and have expected compensation of $E(\pi_t) = \alpha(0) + \beta$, although their realized compensation depends on their exact action. If they do not trade their compensation will be will be $\pi_t = \alpha(0) + \beta$. If they trade ξ and guess correctly on its end of period value their compensation will be $\pi_t = \alpha(\frac{1}{2}) + \beta$ and if they trade ξ but guess incorrectly on its end of period value their compensation will be $\pi_t = \alpha(-\frac{1}{2}) + \beta$.

Investors participate in the market if the asset payoff minus managerial compensation is positive in expectation. The equation below solves for the investor's participation constraint: they will hire a random manager if

$$\gamma(1-\alpha)(\frac{1}{2})(\frac{N\eta}{\gamma+N\eta}) + (1-\gamma)(1-\alpha)(0) - \beta \geq 0$$

$$\rightarrow \gamma(1-\alpha)(\frac{1}{2})(\frac{N\eta}{\gamma+N\eta}) \geq \beta \tag{30}$$

This is not guaranteed if $\beta > 0$ and $\alpha \rightarrow 0$, so investor participation requires an additional assumption. I will make the assumption a little stronger than the final equation above to ensure that investors will participate even if bad managers are randomly trading on ν. In that case (i) the spread decreases, but (ii) bad managers incur losses equal to half of the spread with certainty, even though they only guess the correct asset value half of the time. The cost of bad managers paying the spread outweighs the benefit of good managers having a slightly smaller spread, so the investor is slightly worse off with the bad manager guessing at the value of ν and requires a slightly stronger assumption to ensure participation. The assumptions on parameters that guarantee investor participation even if bad managers are trading ν are summarized in Assumption 2.

Assumption 2 $\gamma(1+\alpha)(\frac{1}{2})(\frac{N\eta}{1+N\eta}) \geq \beta$

The only difference between the participation constraint in the separating equilibrium where bad managers do not trade ν and a pooling equilibrium where bad managers do trade ν is the denominator of the last term on the left hand side. In a separating equilibrium the denominator is $\gamma + N\eta$, while in a pooling equilibrium the denominator is $1 + N\eta$. The assumption for the participation constraint is slightly stronger in the pooling equilibrium, which reflects that the investor's expected welfare is slightly lower.

In the equilibrium without career concerns good managers will perform better than bad managers in expectation. However, a lucky bad manager will outperform the good manager:

$$\chi_{t,lucky} = \frac{1}{2} > \frac{1}{2}\frac{N\eta}{\gamma+N\eta} = \chi_{t,good} \tag{31}$$

The initial definition of gambling provided in the introduction and discussed thoroughly in the previous chapter of this thesis is bad managers intentionally increasing the variance of their portfolio returns. This happens most intuitively through their leverage or selection of assets with higher idiosyncratic volatility or higher factor exposures. However, in the model, the same outcome - lucky bad managers outperform good managers - has arisen even though the payoffs of the two assets are identical. Here, the outperformance comes from the small penalty the good manager incurs when trading ν against a market maker who knows there is the possibility of being matched with an informed agent. The good manager always has to pay half of the spread demanded by the market maker. In contrast, by trading the noise asset, the bad manager does not pay any spread and outperforms the good manager with probability 0.5. Despite the assets having the same variance, the gambling bad manager is able to achieve more extreme positive payoffs. The model has illuminated another potential channel which allows lucky bad managers to outperform good managers: by not trading an asset where good managers have additional information, they can save money on market makers' spreads.

In the following section I will consider the equilibrium when a rational, fully informed investor makes an optimal re-hiring rule instead of re-hiring managers exogenously.

2.2.2 Equilibrium with no credible ex-ante commitment

When investors make a re-hiring decision at the beginning of the second period, they can observe the payoffs the manager earned in the first period. Since the good manager has higher expected payoffs than the bad manager, they would like to set a re-hiring rule that allows them to only re-hire a good manager. However, since the bad manager can imitate the good manager by guessing correctly in the market for ν, it's possible that the investor cannot perfectly discern managerial quality over short horizons.

The best way for an investor to discern if the manager is good or bad is to only re-hire a manager only if their payoff was exactly $\chi_1 = \frac{1}{2} - \frac{1}{2}(spread_\nu)$. That is, the investor can decide to only re-hire a manager that traded correctly in the market for ν. If the bad manager is acting identically to the equilibrium without career concerns they will be trading ξ or not trading. If the investor then observes a payoff $\chi_1 = \frac{1}{2} - \frac{1}{2}(spread_\nu)$ it reveals the manager is good with probability 1. Even if the bad manager changes their behavior and trades ν in an attempt to imitate the good manager the probability the manager is good conditional on seeing a correct trade in ν is $\frac{\gamma}{\gamma+0.5(1-\gamma)} = \frac{2}{1+\gamma}\gamma > \gamma$. Since the bad manager is successful with gambling in ν only half of the time, successfully trading in ν is a credible signal of managerial quality. Without commitment to a re-hiring rule that has profitable ex-post deviations, the investor optimally chooses this rule and only re-hires a manager who successfully trades ν.

The next section will examine the possibility of the rational investor designing a re-hiring rule at the beginning of period 1 that will allow them to achieve higher welfare by allowing bad managers to gamble instead of imitating good managers. However, without ex-ante, credible commitment to not profitably deviate from a re-hiring rule, fund managers of either quality know that any manager not earning $\chi_1 = \frac{1}{2} - \frac{1}{2}(spread_\nu)$ is revealed as a bad manager and would be fired.

The following assumption clarifies the bounds on the number of noise trades occurring in every period to justify the investor's rule. The importance of the assumption will be discussed in the paragraphs following Proposition 2.

Assumption 3 $N\eta > \frac{1+\gamma}{\gamma}$

Under Assumption 3, Proposition 2 identifies the unique equilibrium with rational investors who cannot commit an a re-hiring rule that has profitable deviations.

Proposition 2 *If the investor is unable to credibly commit to a re-hiring rule with profitable deviations, they will optimally re-hire any manager who trades correctly in ν. Under this rule, good managers trade ν using their informative signal and bad managers trade ν randomly.*

Good managers, as always, trade profitably in ν. Bad managers have the choice of continuing to trade in ξ or not trading, where they will be fired with probability 1, or switching to randomly trading to ν. If they trade randomly in ν they will be re-hired on a successful guess on the asset's end of period value, which occurs with probability 0.5. The downside of trading ν is the manager loses half of the market maker's spread with certainty, and their expected compensation in the first period becomes

$$E(\pi_1) = \alpha(-\frac{1}{2}\frac{\gamma}{\gamma + N\eta}) + \beta \tag{32}$$

In the second period bad managers return to randomly trading ξ or not trading, earning an expected $E(\pi_2) = \alpha(0) + \beta$. This occurs with probability 0.5, since there is only a 0.5 chance the bad manager guesses correctly in ν in the first period. Their total expected compensation from trading ν randomly in period 1 is then $E(\pi) = \alpha(-\frac{1}{2}\frac{\gamma}{\gamma+N\eta}) + \beta + \frac{1}{2}\beta$. In the second period, if they were lucky in the first period trade, the bad manager returns to trading ξ or not trading and earn β in expectation. This is compared with $E(\pi) = (0)\alpha + \beta = \beta$ for trading ξ (or not trading) and revealing their type with probability 1. To summarize, they will switch to trading ν if

$$\alpha(\frac{1}{2}\frac{\gamma}{\gamma + N\eta}) \leq \frac{1}{2}\beta \tag{33}$$

Which holds when $\beta > 0$ and $\alpha \to 0$. A 50% chance of earning their fixed compensation β in $t = 2$ outweighs the loss of paying half the spread with certainty.

When investors condition re-hiring on a manager earning the exact payoff of the good manager, bad managers change their behavior and attempt costly imitation.

If this rule is implemented, the bad manager changes their action and moves to a pooling rather than separating equilibrium. The investor deciding if this rule is optimal must weigh the cost incurred by introducing the rule, which is any bad manager trading ν and paying half the spread with certainty in $t = 1$, against the benefit of introducing the rule, which is increasing the probability of having a good type manager in $t = 2$. The cost of this rule depends on both the spread (set by the number of noise traders) as well as the ex-ante probability of hiring a good versus bad manager. If a good manager is hired with too little probability or the spread is too large, it's not worth it for the investor to incentivize bad managers to engage in costly imitation of the exact actions of the good manager. Assumption 3 ensures that there are enough noise traders such that the spread is not too large and the probability of hiring a good manager is not too low, and the investor's welfare improves by adopting the rule to only re-hire managers trading correctly in ν. For example, in the simple case where $\gamma = 0.5$ - good and bad managers are hired with equal probability - there must be $N\eta = 3$ noise trades for the investor to improve welfare by switching to this rule. The amount of noise traders required to improve investor welfare decreases with larger probabilities of γ.

Market makers adjust prices to account for the presence of an additional noise trader (the bad manager) and set prices and spreads as in Equations 27, 28, and 29. The investor's expected profit in the first period is $\gamma(1-\alpha)\frac{1}{2}\frac{N\eta}{1+N\eta} - \beta$.

When career concerns are present for a fully informed, rational investor, bad managers switch to attempting to exactly imitate the good managers. They no longer outperform the good manager when they receive a lucky realization of asset value. By definition anything that is unique in the public information of bad quality managers, like earning a negative payoff or a positive payoff that is too high, instantly reveals their type and causes them to be fired.

One interesting feature of this equilibrium is it relies on investors being able to perfectly discern *holdings* from their observation of payoffs, which is difficult in reality (Dasgupta and Piacentino (2015)), especially when funds window dress their holdings (Kacperczyk et al. (2008), Agarwal et al. (2014)). The rule examined in this section is a theoretical ideal but impractical: when an investor in reality observes returns they do not know how the fund actually traded to obtain their return, as holdings data are not provided continuously. For example, a fund may fully buy then sell out of a position within the three month gap between holdings disclosures.

However, even with investors being able to perfectly discern holdings from returns, investors may still optimally commit to a simpler rule that allows the bad manager to reveal their type. They can improve their welfare if they can design and commit to a rule that allows for the separating instead of the pooling equilibrium in each period, while still somehow firing the bad manager with the same probability. The next section discusses how the optimal re-hiring decision changes if investors are able to credibly commit ex-ante to a re-hiring rule with profitable deviations.

2.2.3 Equilibrium with credible ex-ante commitment

Suppose the investor can commit to any re-hiring rule before hiring the fund manager, even if the rule had profitable ex-post deviations. If so, a fully rational investor may want to consider a rule that allows the bad manager to gamble. The intuition is simple: attempting to imitate the good manager is costly for the bad manager as well as for the investor, since bad managers pay half of the spread with certainty for every guess in the market for ν. If the investor was able to re-hire the bad manager for a successful gamble in ξ instead of ν, their welfare could improve: they would still fire the bad manager with the same probability but they would not incur a fixed cost each time the bad manager trades, since the bid-ask spread of ξ is zero. The investor's commitment to the rule would have to be credible, however, as the investor would have an incentive to deviate from the rule any time they observed a manager with a payoff $\chi_1 \notin \frac{1}{2} - \frac{1}{2}(s_{\nu,t})$, as any payoff $\chi_1 \notin \frac{1}{2} - \frac{1}{2}(s_{\nu,t})$ cannot be realized by a good manager. In other words, the investor would have to credibly commit to re-hiring a lucky bad manager, even though the bad manager has revealed their type.

Proposition 3 summarizes the optimal rule for the fully rational investor which allows bad managers to be re-hired even after they reveal their type.

Proposition 3 *If the investor is able to credibly commit to a re-hiring rule with profitable deviations, they will optimally re-hire any manager with a positive payoff realization. Under this rule, good managers profitably trade ν while bad managers trade ξ randomly.*

Identically to previously studied hiring rules, the good manager trades profitably in ν. They earn a payoff $\chi_{1,g} = \frac{1}{2} - \frac{1}{2}(s_{\nu,t})$. The good manager is re-hired with probability 1 since their payoff is always positive. Bad managers know they must trade in order to obtain a positive payoff and be re-hired but unlike the previous section when investors only re-hired managers with the *exact* payoff of a good manager the bad manager can now trade randomly in ξ or ν. They choose to trade in ξ since the noise asset has no spread: their payoff is $\chi_{1,b} \in \{-\frac{1}{2}, \frac{1}{2}\}$, each occurring with probability 0.5. Their expected payoff is zero and their expected two period compensation is $E(\pi) = \alpha(0) + 1.5\beta$. This is higher than their expected total compensation if they traded randomly in ν, which would be $E(\pi) = \alpha(-\frac{1}{2}(s_{\nu,t})) + 1.5\beta$. The market maker sets prices and spreads as defined in Equations 24, 25, and 26. This is an identical equilibrium to the model without career concerns. The investor's expected profit in period 1 is $\gamma(1-\alpha)\frac{1}{2}\frac{N\eta}{\gamma+N\eta} - \beta > \gamma(1-\alpha)\frac{1}{2}\frac{N\eta}{1+N\eta} - \beta$, where the right hand side of the equation is the investor's expected profit in the pooling equilibrium generated by the previous rule of re-hiring only a manager who trades profitably in ν. The improvement to the investor's welfare takes into account the fact that good managers now pay a slightly higher spread when trading ν since market makers know there is a lower probability of encountering a noise trader (the bad manager) in ν. Overall, the benefit of steering the bad manager away from costly imitation outweighs the small additional cost in the good manager's spread.

The surprising result of this equilibrium is that investors sometimes knowingly re-hire bad managers who earn a payoff that is too high to come from a good manager. As a reminder, if the bad manager trades ξ successfully, they earn a payoff $\chi_t = \frac{1}{2} > \frac{1}{2} - \frac{1}{2}(s_{\nu,t})$, where the right hand side of the inequality is the payoff of a good manager. However, from an ex-ante perspective, this is efficient and welfare improving. Regardless of if the bad manager is trading ν or ξ they have a positive payoff with probability 0.5. In the pooling equilibrium where the bad manager trades ν, the bad manager and investor split a fixed cost equal to half the spread every time the bad manager attempts to imitate the exact trade of a good manager. According to their compensation contract α of this cost is borne by the manager and $1 - \alpha$ by the investor. Both parties would be better off if they could avoid this cost. By committing to the rule of re-hiring any manager with a positive payoff the investor makes both themselves and the bad manager better off.

The fully optimal rational rule in this setting looks like a rule a much simpler investor would have made, perhaps from a lack of attention: hire any manager with a positive return. Investors flowing into any mutual fund with a positive raw return has been thoroughly documented in empirical literature: see, e.g. Chevalier and Ellison (1997), Cuthbertson et al. (2010). Managers who earn positive raw returns are supported with inflows instead of being fired, even if the return is suspiciously high. These positive flows hold even if the flows eventually reverse over medium term horizons, as is shown in my first chapter, where flows to extreme right tail performers is very positive in the month following the return but on average negative in the twelve months following the extreme return. However, the results of this model shows that re-hiring or increasing the compensation of managers who earn positive returns is not inconsistent with rationality. Rational investors design a rule that fires the bad manager with probability 0.5 but does not incentive the bad manager to engage in costly imitation.

2.3 Conclusion

This paper investigates different optimal re-hiring rules of a rational investor in the mutual fund industry. The investor first hires a random mutual fund manager who can be good or bad quality. The investor then chooses a re-hiring rule. They would ideally like to only re-hire the good manager. However, the bad manager can attempt to imitate the exact trades of a good manager, which is costly for both the bad manager and the investor. A bad manager guessing correctly in the value of an expertise asset is indistinguishable from a good manager, although they only succeed in their guess with probability 0.5.

If the investor cannot commit to a re-hiring rule with profitable deviations they will only re-hire a manager with the exact trades of a good manager. The benefit of firing a bad manager and replacing them with a new manager in the following period outweighs the cost of the bad manager's imitation. Under this re-hiring rule the bad manager optimally attempts to imitate the exact trades of a good manager, despite the cost of imitation.

However, if investors are able to credibly commit to a re-hiring rule even though it has profitable ex-post deviations, it is rational and welfare maximizing to sometimes re-hire bad managers who reveal their type. The investor optimally chooses a simple rule to re-hire any manager who earns a positive payoff. Bad managers change from attempting to exactly imitate the good manager (trading randomly in the expertise asset) to trading randomly in the noise asset. By allowing bad managers to gamble they no longer incur a fixed cost of imitating the good manager, which is partially paid by the investor through the manager's compensation contract. In this equilibrium good managers outperform bad managers in expectation but a lucky bad manager earns a higher payoff than the good manager. Identically to the previous rule where managers were only re-hired by making the exact trade of a good manager, the bad manager is revealed with probability 0.5. Both the bad manager and the investor have higher welfare in this separating equilibrium relative to the pooling equilibrium where the bad manager engages in costly imitation.

The simple rule of re-hiring a mutual fund manager with a positive return (or investors flowing into any mutual fund with a positive return) is well documented in the empirical literature, and is often taken as a sign of irrationality. Any defense of the rationality of flows has come from an attempt to link flows to positive future performance. This study illuminates a new dimension in how simple flows may not be inconsistent with irrationality: by allowing bad managers to gamble and avoid costly imitation of the good manager's exact trades, the investor improves their own welfare.

2.4 Appendix: Proofs for Optimal Re-Hiring Rules of Mutual Fund Managers

2.4.1 Proof of Proposition 1: equilibrium with exogenous re-hiring

Without career concerns, each manager maximizes their within-period expected payoff χ_t. Since the fixed portion of their compensation β by definition is not dependent on performance, optimal managerial behavior is characterized by maximizing α. For the good manager, they trade ν:

$$\chi_{g,\nu} = \frac{1}{2} - \frac{1}{2} s_{\nu,t}$$

The good manager's payoff χ_t is greater than zero when $s_{\nu,t} < 1$, which is guaranteed by the presence of noise traders.

The bad manager optimally trades ξ or doesn't trade to maximize their expected variable compensation α. Their payoff is

$$E(\chi_{b,\xi}) = 0.5(\frac{1}{2}) - 0.5(\frac{1}{2}) = 0$$

$$\chi_{b,\emptyset} = 0$$

For trading ξ or not trading, respectively. If they tried to trade ν their expected payoff would be

$$E(\chi_{b,\nu}) = -\frac{1}{2} s_{\nu,t} < 0$$

The inequality holds if the spread is greater than zero, which is guaranteed because of the possibility of an informed agent in the market. the bad manager is worse off trading ν than trading ξ or not trading. If they trade ν they incur half of the spread as a cost with certainty but have an expected benefit of zero as the trade is randomly without information on the end of period value of ν.

Market makers believe the good manager trades ν, the bad manager trades ξ or does not trade, and sets prices and spreads for ν as

$$p^A_{\nu,t} = \frac{\gamma}{\gamma + N\eta}(1) + \frac{N\eta}{\gamma + N\eta} * \frac{1}{2} = \frac{1}{2}\frac{2\gamma + N\eta}{\gamma + N\eta} > \frac{1}{2}$$

$$p^B_{\nu,t} = \frac{\gamma}{\gamma + N\eta}(0) + \frac{N\eta}{\gamma + N\eta} * \frac{1}{2} = \frac{1}{2}\frac{N\eta}{\gamma + N\eta} < \frac{1}{2}$$

$$p^A_{\nu,t} - p^B_{\nu,t} = \frac{1}{2}\frac{2\gamma + N\eta}{\gamma + N\eta} - \frac{1}{2}\frac{N\eta}{\gamma + N\eta} = \frac{\gamma}{\gamma + N\eta}$$

Since the spread on ν is between 0 and 1 neither agent changes their actions. Finally, market makers set the prices of ξ as

$$p^A_{\xi,t} = p^B_{\xi,t} = \frac{1}{2}$$

$$p^A_{\xi,t} - p^B_{\xi,t} = 0$$

The investor participates if

$$\gamma(1-\alpha)(\frac{1}{2})(\frac{N\eta}{\gamma + N\eta}) \geq \beta$$

Which holds with Assumption 2.

2.4.2 Proof of Proposition 2: equilibrium without ex-ante commitment

Investors cannot commit to a rule to credibly re-hire a bad manager after they reveal their type, so their two viable choices of re-hiring rules to maximize welfare are (1) an exogenous re-hiring rule and (2) only re-hiring a manager who trades exactly like a good manager (trading correctly in ν). Since the bad manager loses money for the investor (they cost β but generate a maximum expected benefit of 0) no welfare maximizing rule would be chosen to attempt to fire the good manager but re-hire the bad manager.

Suppose the investor puts into place the rule to re-hire a manager has the exact trade of a good manager. Relative to the exogenous re-hiring rule, the bad manager will change their action and trade randomly in ν if

$$\frac{1}{2}\beta > \alpha(-\frac{1}{2}(s_{\nu,t}))$$

The left hand side is their benefit of imitation: they are re-hired with probably $\frac{1}{2}$ on a correct guess in the end of period value of ν, earning β. The right hand side is the cost of imitation: they pay a cost of half the spread with certainty but gain no expected benefit from their guess in ν. Since $\beta > 0$ and $\alpha \to 0$, this inequality holds. The bad manager will change their action in $t = 1$. In the final period both agents will return to their within-period optimum, with the good manager trading ν profitably and the bad manager trading ξ randomly or not trading.

The benefit to the investor to changing rules favoring the good manager is that they raise the probability a manager is good in period 2. Specifically, the probability an agent is good upon seeing a successful trade in ν is

$$p(g|\nu_+) = \frac{2\gamma}{1+\gamma}$$

This increases the probability of good relative to hiring a random manager by

$$\Delta p(g) = \frac{2\gamma}{1+\gamma} - \gamma = \frac{\gamma(1-\gamma)}{1+\gamma}$$

The cost to the investor of changing rules is that the bad manager is now trading ν, which incurs a deadweight loss. The deadweight loss is equal to half of the spread.

In order to show that the investor's welfare improves with the re-hiring rule I will show that the expected benefit of switching from exogenous hiring rules outweighs the expected cost. Since the market maker's exact pricing is complicated - when an agent trades correctly in ν they update prices and spreads - I will first assume the market maker charges the maximum feasible spread to simplify the problem. These prices would be the optimal prices of the market maker is the probability the manager is good is equal to one. Since the *benefit* of changing re-hiring rules decreases in the spread (as the good manager also pays half the cost of the spread for each trade) and the *cost* of changing re-hiring rules increases in the spread, if changing re-hiring rules increases welfare under the maximum possible spread, it will also increase welfare for any other choice of spread.

The maximum feasible spread and the accompanying bid and ask prices are as follows:

$$p^A_{\nu,max} = \frac{1}{1+N\eta}(1) + \frac{N\eta}{1+N\eta}(\frac{1}{2}) = \frac{2+N\eta}{2+2N\eta}$$

$$p^B_{\nu,min} = \frac{1}{1+N\eta}(0) + \frac{N\eta}{1+N\eta}(\frac{1}{2}) = \frac{N\eta}{2+2N\eta}$$

$$s_{\nu,max} = \frac{1}{1+N\eta}$$

The expected benefit of changing re-hiring rules is then

$$\Delta p(g)(\chi_{g,2}) = (\frac{\gamma(1-\gamma)}{1+\gamma})(\frac{1}{2}\frac{N\eta}{1+N\eta})$$

An the expected cost is

$$p(b)(\frac{1}{2}s_{\nu,max}) = (1-\gamma)\frac{1}{2}\frac{1}{1+N\eta}$$

Comparing the two, changing re-hiring rules increases welfare if

$$(1-\gamma)\frac{1}{2}\frac{1}{1+N\eta} < (\frac{\gamma(1-\gamma)}{1+\gamma})(\frac{1}{2}\frac{N\eta}{1+N\eta})$$

$$\rightarrow \frac{1+\gamma}{\gamma} < N\eta$$

Which holds under Assumption 3.

Since the investor was participating in the market with exogenous re-hiring rules and their welfare has increased, they will also participate in the market while re-hiring any manager who trades ν successfully.

2.4.3 Proof of proposition 3: equilibrium with credible ex-ante commitment

For this proof, I will establish that the rule of hiring any manager with a positive return increases welfare over the previous rule, which is re-hiring any manager who trades correctly in ν.

Under the new re-hiring rule the good manager, as always, profitably trades on ν. The bad manager will trade ξ since

$$\alpha(-\frac{1}{2}(s_{\nu,t})) + \frac{1}{2}\beta < \alpha(0) + \frac{1}{2}\beta$$

The left hand side of the equation is their expected payoff for trading ν randomly in period one, while the right hand side is the expected payoff for trading ξ randomly in period one. In both cases they have a probability of 0.5 of being re-hired on a successful guess under the rule that any manager with a positive return is re-hired, giving them a 50% chance of earning β in $t = 2$. Trading ν negatively impacts their compensation through the variable term α because they pay half the spread as a cost with certainty but the benefit expected benefit of a random guess in ν is equal to zero.

Since under either rule the investor fires the bad manager with probability 0.5, their expected welfare in $t = 2$ is identical under either rule. Relative to the pooling equilibrium, in $t = 1$, the investor's welfare has increased in expectation by

$$(1-\gamma)(1-\alpha)(\frac{1}{2}(s_{\nu,t})) > 0$$

by incentivizing the separating equilibrium where the bad manager pays ξ. The equation is intuitive to interpret: if a bad manager is chosen ($p = 1 - \gamma$), the investor no longer shares via the compensation contract $(1 - \alpha)$ half the market maker's spread $(\frac{1}{2}(s_{\nu,t}))$.

Since the separating equilibrium with the bad manager trading ξ is the welfare maximizing equilibrium for the investor within any individual period, a better rule only exists if it can increase the probability of the bad manager being fired. However, no rule exists that would hire the bad manager with probability greater than 0.5 under their current compensation contract. The bad manager would always choose to attempt to imitate the good manager over revealing their type and being fired with probability 1 since

$$\frac{1}{2}\beta > \alpha(-\frac{1}{2}(\frac{1}{1+N\eta}))$$

Under the assumption that $\beta > 0$ and $\alpha \to 0$.

To summarize, re-hiring any manager with a positive return allows the investor to incentivize the within-period welfare maximizing equilibrium. It also has the highest possible probability of firing the bad manager. Therefore, no better rule exists that increases the investor's welfare.

3 Entrepreneur Participation with Limited Liability and Bankruptcy

3.1 Introduction

Limited liability has been recognized as a crucial tool in incentivizing risk averse entrepreneurs to begin new firms. By protecting entrepreneurs from the downside risk of firm failure, it encourages entrepreneurship and new firm creation. This is especially important for firms that are innovative and face high uncertainty about the profitability of their products. For example, *The Economist*, recognizing the role of limited liability in fuelling the Industrial Revolution, wrote: "The economic historian of the future may assign to the nameless inventor of the principle of limited liability, as applied to trading corporations, a place of honour with Watt and Stephenson, and other pioneers of the Industrial Revolution. The genius of these men produced the means by which man's command of natural resources was multiplied many times over; the limited liability company the means by which huge aggregations of capital required to give effect to their discoveries were collected, organized and efficiently administered."[9]

Despite its widely acknowledged benefits, limited liability firms worldwide exhibit high variation in entry costs. This can be in the form of direct costs for registration fees or indirect costs, for example, high minimum capital requirements. In the UK, a limited liability company can be set up for £12 (or £30 for expedited processing) and no minimum capital: any amount about £0 is sufficient. In contrast, in Austria, the minimum capital requirement for a limited liability company (the Gesellschaft mit beschränkter Haftung, or GmbH) is €10,000. Before December 2023, it was €35,000. The fees for registration of a GmbH in Austria vary depending on the size of capital initially committed.

Early opponents of limited liability were worried about fraud. High barriers to entry, or the outright restriction of the limited liability firm type, was justified by the claim that free access to limited liability would lead to large amounts of fraudulent entrepreneurs tricking investors and consumers into giving them money (Halpern et al. (1980)). This hypothesis has been tested by several by several modern finance papers. Djankov et al. (2002) find that higher barriers to entry for limited liability firm types are associated with high levels of corruption and a larger size of the informal economy, but not with high quality products, fewer externalities, or greater competition. Klapper et al. (2006) find that higher barriers to entry leads to lower firm creation and lower growth of incumbent firms. Both suggest that high barriers to entry have more to do with rent protection than consumer or creditor protection.

[9] The Economist, 18 December 1926. Taken from Halpern et al. (1980), who in turn cites a 1936 article by B. C. Hunt.

A country looking to increase entrepreneur participation may then conclude that it should lower barriers to entry; by opening access to the limited liability corporate form, entrepreneurs will be enticed to participate in firm creation. However, entrepreneur risk aversion can be catered to through a variety of tools. Another common tool to encourage participation is generous bankruptcy. Founding an unlimited liability firm, and being protected by bankruptcy if the firm fails, should also encourage entrepreneur participation. Fan and White (2003) finds this to be the case; states with more generous bankruptcy exemptions have higher levels of workers classifed as self-employed. But Berkowitz and White (2002) finds that generous bankruptcy is also associated with credit constraints among entrepreneurs. An entrepreneur in a state with more generous bankruptcy is less likely to be able to get a loan, and conditional on receiving a loan, faces a higher interest rate. Since the 'barriers to entry' papers do not explicitly test for credit constraints, should we expect this to be true from making limited liability cheaper to access as well? Limited liability may also allow for poorer quality entrepreneurs to participate in firm creation, leading to similar credit constraints. Is there a meaningful difference between making bankruptcy more generous and making limited liability cheaper to access?

By examining the choice of contracts offered by limited liability versus unlimited liability and generous personal bankruptcy, this paper provides guidance on why bankruptcy leads to credit constraints while lower barriers to entry for limited liability firms does not. I consider a financing problem with an entrepreneur subject to a moral hazard problem requiring skin in the game, i.e. requiring wealth to be lost when the firm fails. Wealth is fully liquid, and there is no asymmetric information. The entrepreneur can finance using equity or debt, and can choose limited or unlimited liability firm types, e.g. a LLC vs a sole proprietorship. The first section of the paper addresses 'synthetic' regimes. Synthetic limited liability means creating limited liability consumption outcomes for an unlimited liability firm, and synthetic limited liability is symmetrically defined. Synthetic regimes relate to how easy it is to make 'homemade' limited or unlimited liability, when only given access to the opposite regime. The results show that synthetic limited liability is costly, while synthetic unlimited liability is not. Then, limited liability admits the largest choice of entrepreneur repayments, even though it is a contractual restriction on punishment. Analyzing the choice of repayments under each regime shows that unlimited liability can be thought of as a mandate for maximum punishment rather than an absence of restriction on punishment. This holds under both liquid and illiquid wealth.

The observation that synthetic unlimited liability - unlimited liability consumption outcomes with a limited liability firm - is costless has implications for the entrepreneur's moral hazard. The argument that open access to limited liability allows for bad quality and fraudulent entrepreneurs has proven unfounded empirically (Djankov et al. (2002), Klapper et al. (2006)). The results of the model explain why poor quality entrepreneurs are able to flourish with cheap access to limited liability. Any moral hazard problem which can be solved by the entrepreneur having a large enough stake in the firm can can be solved equally well by limited and unlimited liability. Ex-post, limited liability can lead to a number of moral hazard issues, as a large literature documents. But ex-ante, there is no inherent problem with the limited liability firm type. Moral hazard problems of this type cannot be a justification for restricting access to limited liability.

The combination of bankruptcy and unlimited liability is redundant with limited liability when wealth is fully liquid for an entrepreneur with sufficiently low outside wealth. If limited liability forms are made costly by governments, as they are in many countries, generous bankruptcy can help poorer entrepreneurs access limited liability risk sharing benefits even when they incorporate under the unlimited liability corporate form. However, when wealth is illiquid, more generous bankruptcy is a strict reduction of the possible choice of contracts under both limited and unlimited liability. When personal bankruptcy is made more generous, some state contingent repayments are no longer possible: an entrepreneur can always declare bankruptcy if the creditor attempts to collect outside wealth protected by bankruptcy laws. Then, the entrepreneur must liquidate their wealth ex-ante and pay it into the firm to solve the moral hazard problem, which is more costly than only liquidating wealth if the firm fails. If wealth is illiquid, more generous personal bankruptcy leads to larger deadweight losses of liquidation for entrepreneurs: they can no longer credibly repay creditors from outside wealth (by, for example, pledging collateral protected by bankruptcy laws) once their firm has failed.

In a country with barriers to entry to limited liability firms, bankruptcy encourages some entrepreneurs to participate by allowing cheap access to limited liability consumption outcomes. With illiquid entreprneuer wealth, it will also decrease the amount entrepreneurs are able to pledge as repayment when the firm fails. The results of the model explain why prior empirical studies have found that generous bankruptcy increases entrepreneur participation but results in credit constraints while lowering the cost of limited liability firm types has a strictly positive impact on entrepreneurship.

3.1.1 Literature Review

This paper draws from two main strands of literature. First, the impacts of limited liability and the agency problems it creates has a long history in legal and financial thought. Theoretical works include Jensen and Meckling (1976), who first outlined many common agency problems inherent in limited liability firm structures. Easterbrook and Fischel (1985) examine liability from a legal perspective and considers its uses as well as the uses of courts ‘piercing the veil’, i.e. stripping limited liability protections for owners or managers of a firm. The closest model to this paper is Winton (1993), which models optimal choice of shareholder liability under different assumptions for the verifiability and liquidity of wealth. In comparison, I discuss optimal choice of liability for entrepreneurs in modern times with the ability to pledge assets as collateral, and find that for these agents the limited liability firm type dominates unlimited liability.

Limited liability has long been the purview of theoretical works due to clear selection biases in firms choosing their legal form. However, recent empirical literature has managed gain insight on the impacts of limited liability using natural experiments. Koudijs and Salisbury (2020) uses a natural experiment in marital property laws among the US South the 1840s to find how liability protections impact entrepreneurial activity. They find that some liability protection increases entrepreneurial investments, but that too much is bad for moral hazard. However, the limited liability exemptions they study are more like bankruptcy protections than modern limited liability, since they change what wealth can be seized in the event of a firm’s default: in other words, the laws regulate state contingent repayment of entrepreneurs when their firms fail. This paper is able to make the distinction between the effect of generous bankruptcy versus easy access to limited liability, and explain why the authors find mixed results on entrepreneur participation.

The most related strand of literature to my results comes from the study of entrepreneur incentives, barriers to entry for limited liability firms, and bankruptcy policies. Gottlieb et al. (2022) study the link between entrepreneurship and risk aversion and find that a longer window of maternity leave for a new mother leads to more entrepreneurship. Longer maternity leave means a mother has a longer time to potentially explore entrepreneurial activities while the return to her old job is protected by law. Gropp et al. (1997) find that bankruptcy can lead to credit constraints for low income households; the probability of being denied a loan is higher, and conditional on receiving a loan the interest rate is higher, than comparable states with lower bankruptcy thresholds. Fan and White (2003) finds that states with more generous bankruptcy exemptions have higher levels of workers qualified as self-employed. Berkowitz and White (2002) investigates both participation and credit constraints in entrepreneurs. For states with more generous bankruptcy, there are higher levels of entrepreneurship but credit constraints among entrepreneurs. Entrepreneurs in generous bankruptcy states have a lower chance of receiving a loan, and conditional on receiving a loan, have higher interest rates.

In contrast, Djankov et al. (2002) and Klapper et al. (2006) investigate barriers to entry for limited liability firms and find that lowering barriers to entry increases entrepreneurial activity but does not result in lower quality firms. Both papers reject the conclusion that governments need high barriers to entry to protect creditors and consumers from poor quality entrepreneurs. The results of these two papers will be discussed in more detail in Section 4. My model explains why papers in the previous paragraph studying bankruptcy find mixed results on entrepreneurship while studies of high barriers to entry find that entrepreneurship strictly declines when limited liability firms are costly to form.

The remainder of the paper is laid out as follows. Section 2 introduces a model of entrepreneur financing with liquid wealth. Section 3 considers illiquid wealth. Section 4 applies the model to investigate making bankruptcy more generous compared to making limited liability firms cheaper to create. Section 5 discusses the prevalence of unlimited liability firms, and Section 6 concludes.

3.2 A model of entrepreneur financing

Consider a simple financing problem between a single risk neutral entrepreneur and a continuum of competitive, risk neutral creditors. The model will build on Holmstrom and Tirole (1997). There are three discrete time periods, $t \in \{0, 1, 2\}$. In $t = 0$ the entrepreneur can start a firm, which costs I, and provides some output in $t = 2$. For simplicity, there is no time discounting. The entrepreneur has starting wealth $W_i < I$ and can choose to pay $A_i \leq W_i$ of their personal wealth towards the financing cost. The entrepreneur must contract with outside creditors for the remaining $I - A_i$. The firm produces cashflows $X_2 \in \{r, R\}$, with $0 < r < R$.

The entrepreneur can choose between debt and equity financing and between limited and unlimited liability regimes. With an equity contract α of the firm's cashflows accrue to the entrepreneur while $(1 - \alpha)$ goes to shareholders. A debt contract under limited liability is negotiated with face value F such that creditors are paid until payments satisfy F or until the firm is out of money. In model notation, creditors receive $\min\{F, X_2\}$. In an unlimited liability regime the creditors can additionally access the entrepreneur's external wealth to repay their claim: they receive $\min\{F, X_2 + (W_i - A_i)\}$.

The entrepreneur makes an unobservable effort choice at $t = 1$: they can work or shirk. If they work the probability of the high payoff is p_H. If they shirk the probability of the high payoff is $p_L < p_H$. Additionally, shirking provides a private, nontransferable benefit B. Assumption 1 establishes that the firm is positive NPV when the entrepreneur works and negative NPV when the entrepreneur shirks, even when the value of the private benefit is included.

Assumption 1 $p_H R + (1 - p_H)r - I > 0$, $p_L R + (1 - p_L)r - I + B < 0$

The entrepreneur raises financing from a continuum of sufficiently wealthy, risk neutral, competitive creditors. Creditors have full information and must break even in expectation: in particular, they can observe the moral hazard problem facing the entrepreneur. Since the project is negative NPV if the entrepreneur shirks the project will only be financed if the entrepreneur works: the break-even constraint of the creditors will use p_H as the probability of the firm's success. Their break-even constraint if equity finance is used is

$$(1-\alpha)p_H R + (1-\alpha)(1-p_H)r \geq I - A_i \tag{34}$$

The break-even constraint will be solved with equality in equilibrium since creditors are competitive.

To make the problem interesting when debt financing is used I additionally assume that the entrepreneur has too little wealth to allow for risk free debt even with $A_i = W_i$:

Assumption 2 $W_i < \frac{I-r}{1+p_H}$

This ensures that even if $W_i = A_i$ the face value of debt F satisfying the break-even constraint will be greater than r.

The creditor's break-even constraint for debt financing with face value F is mechanically

$$p_H F + (1-p_H)(r + UL) \geq I - A_i \tag{35}$$

Where UL indicates any additional repayment from the entrepreneur's wealth under unlimited liability. In other words, $UL = W_i - A_i$ if the entrepreneur chooses the unlimited liability regime. The debt break-even constraint is solved with equality in equilibrium.

An entrepreneur starting a firm at $t = 0$ must choose A_i, debt or equity financing, and the type of liability regime that solves their moral hazard. The solution of the moral hazard problem depends on the regime: the equations below list the moral hazard problem for equity financing, debt financing with limited liability, and debt financing with unlimited liability, respectively.

$$(\alpha)p_H R + (\alpha)(1-p_H)r \geq (\alpha)p_L R + (\alpha)(1-p_L)r + B \tag{36}$$

$$p_H(R-F) + (W_i - A_i) \geq p_L(R-F) + (W_i - A_i) + B \tag{37}$$

$$p_H(R-F) \geq p_L(R-F) + B \tag{38}$$

In the last equation, note that $W_i < \frac{I-r}{1+p_H}$ implies $F > r + (W_i - A_i)\ \forall A_i \leq W_i$. Since debt is sufficiently risky the entrepreneur consumes zero in states when the firm realizes its low cashflow regardless of their choice of A_i. Any money outside of the firm will be claimed by creditors in the low state realization under unlimited liability.

In order to concisely write the entrepreneur's maximization problem, define generically the entrepreneur's repayment to creditors as R_C^S if the firm succeeds and R_C^F if the firm fails. R_C^S and R_C^F are specified by the type of contract as well as the liability regime as detailed above. The entrepreneur's full problem is then

$$\max_{\{A_i, R_C^S, R_C^F\}} p_H(R - R_C^S) + (1 - p_H)(r - R_C^F) \tag{39}$$

s.t.

$$I - A_i = p_H R_C^S + (1 - p_H) R_C^F \tag{40}$$

$$p_H(R - R_C^S) + (1 - p_H)(r - R_C^F) \geq p_L(R - R_C^S) + (1 - p_L)(r - R_C^F) + B \tag{41}$$

The first equation is the entrepreneur's utility maximization, the second is the break-even constraint of creditors, and the third is the entrepreneur's incentive compatibility constraint. Note that a risk neutral entrepreneur would be indifferent between any feasible form of financing: in expectation creditors always receive $I - A_i$ and the entrepreneur receives $NPV - I$. However, the entrepreneur may not be able to commit to solving the moral hazard for all types of financing. In particular, the entrepreneur's minimum stake in the firm that solves moral hazard A_{min}^{IC} is higher for equity than for debt. The minimum stakes are solved by substituting the break-even constraint of creditors into the incentive compatibility constraint of the entrepreneur. For equity, the entrepreneur must have

$$A_{min}^{IC}(E) = [\frac{B}{\Delta p(R - r)} - 1](r + p_H(R - r)) + I \tag{42}$$

For debt, the entrepreneur must have

$$A_{min}^{IC}(D) = p_H \frac{B}{\Delta p} - (p_H R - I) - (1 - p_H) r \tag{43}$$

With $A_{min}^{IC}(D) < A_{min}^{IC}(E)$. The inequality will be shown with more detail in the Appendix proof of proposition 1.

Before moving on to results, the final assumption ensures the entrepreneur is able to at least afford debt financing:

Assumption 3 $W_i > A_{min}^{IC}(D)$

In the first results of the paper, I consider the entrepreneur's wealth to be fully liquid: W_i is costlessly converted in to A_i. In Section 3.3 I consider illiquid wealth.

3.2.1 Synthetic Regimes

The first step to understanding entrepreneur choice of contract under limited and unlimited liability is by trying to replicate their consumption outcomes. How could limited liability consumption outcomes be reproduced in an unlimited liability firm, and how could unlimited liability consumption outcomes be reproduced in a limited liability firm? I will call these synthetic regimes. Synthetic regimes denote contracts under which consumption outcomes are identical between limited and unlimited liability. I will call synthetic limited liability a firm that incorporated under *unlimited* liability firm but has identical consumption outcomes to a limited liability firm, and synthetic unlimited liability a firm that incorporated under *limited* liability but has identical consumption outcomes to an unlimited liability firm. Synthetic regimes denote replicating consumption outcomes using the opposite form of incorporation.

First, consider synthetic limited liability. Creating limited liability consumption outcomes in an unlimited liability firm is tricky; any debt contract under unlimited liability threatens the entrepreneur's outside wealth, while under limited liability outside wealth is completely protected. The simplest way to create synthetic limited liability with a flexible stake A_i is to avoid debt contracts entirely and only finance with equity; with equity financing, you don't need the limited liability firm type, as payments are independent of liability regime.

Proposition 1 *An entrepreneur incorporating under an unlimited liability firm structure can recreate limited liability consumption outcomes by choosing to finance with equity. However, this is costly; the minimum wealth at which entrepreneurs will be able to overcome their moral hazard problem is higher with equity financing than with debt financing.*

The first part of the proposition follows from the pay structure of equity contracts. Without debt, protecting the entrepreneur's outside wealth from repayments to debt is meaningless. The second part of the proposition comes from a well-established result in corporate finance that equity is worse for incentive compatibility than debt, which I replicate in the appendix under this specific model. With equity financing some of the bad state cashflow always goes to the entrepreneur, weakening their incentive to work.

Synthetic limited liability being costly has implications for the entrepreneur's optimal choice of liability regime. If creating synthetic unlimited liability was also costly, we could conclude that each liability regime has its unique purpose as a structure of financing. However, unlike synthetic limited liability, creating synthetic unlimited liability has no cost.

Proposition 2 *An entrepreneur incorporating under a limited liability regime can costlessly replicate unlimited liability consumption outcomes by choosing* $A_i = W_i$.

With debt financing, limited liability is a protection of outside wealth $W_i - A_i$. The simplest way to create synthetic unlimited liability is to commit the entrepreneur's entire outside wealth and have $W_i - A_i = 0$. With no outside wealth, the form of incorporation is irrelevant. The entrepreneur simply changes their initial stake A_i to equal the total value of their starting wealth.

Evaluating these two propositions tells us that limited liability is the more general form of incorporation. Unlimited liability can then be considered as a special case of limited liability with fully liquid wealth. However, limited liability contracts can additionally take on other values of $A_i < W_i$. This leads to the following result about choice of stake under liability regimes:

Corollary 1 *The possible repayments to creditors allowing the entrepreneur to satisfy moral hazard under limited liability is strictly greater than the possible repayments to creditors under unlimited liability.*

Intuitively, limited liability may seem to be the more restricted set of contracts, since it limits the maximum allowable punishment by the entrepreneur. However, when the stake A_i is flexible and there is costless liquidation of an entrepreneur's initial wealth, the limited liability restriction on punishment is meaningless. An entrepreneur can always increase A_i ex-ante to increase their punishment when the firm fails, until the entrepreneur's wealth is exhausted ($A_i = W_i$). In the proof of this corollary in the appendix I will show that choosing debt financing under unlimited liability is identical to choosing $A_i^* = W_i$, regardless of the actual stake: if the entrepreneur chooses $A_i^* < W_i$ all of their outside wealth is still lost when the firm fails, as if they had chosen $A_i^* = W_i$. However, with limited liability debt financing, the entrepreneur can choose some $A_i^* < W_i$ which satisfies the break-even and incentive compatibility constraints and allows them to have positive consumption when the firm fails.

The synthetic unlimited liability contract representation makes it clear that although unlimited liability has no obvious legal restrictions in specifying repayment terms it's equivalent to making the entrepreneur undergo the *maximum* feasible punishment. Because limited liability can represent any possible contractual punishment for failure (including the maximum) and unlimited liability requires maximum punishment for failure, limited liability is the more flexible corporate form.

Although I have shown that limited liability admits a larger choice of repayments, I have not made any claim on why those contracts would be desirable. Why would an entrepreneur want to choose consumption outcomes with $A_i^* < W_i$? The simplest explanation for preferring a smaller stake is if the entrepreneur was risk averse instead of risk neutral. In a firm that uses debt financing, borrows from risk neutral creditors, and incorporates under limited liability, increasing A_i^* beyond the minimum required by incentive compatibility is a mean-preserving spread on a contract with the minimum stake admissible under the moral hazard. Any risk neutral entrepreneur will optimally choose $A_i^* = A_{min}^{IC} < W_i$ for the minimum stake A_{min}^{IC} corresponding to their choice of liability regime and type of financing.

3.2.2 Liability and Moral Hazard

A large body of literature documents that limited liability, once in place, causes moral hazard problems in contracting by separating the payoff of the protected party from the payoff of the firm. These problems go by a variety of names such as effort provision, risk shifting, or tunneling. The cure to all of these problems is to increase the skin in the game of the agent so they bear more of the downside risk of whatever venture they are in charge of. This leads to the impression that limited liability is the cause of all these moral hazard problems. However, there is nothing inherent about the form of liability to any of these issues. Limited liability is compatible with any level of punishment for failure, including the maximum imposed by unlimited liability.

Proposition 3 *For any moral hazard problem which can be solved by requiring the entrepreneur to have a large enough stake in the firm, the problem can be solved equally as well under limited liability as unlimited liability.*

Think again about an entrepreneur attempting to start a firm. The ultimate cure to moral hazard problems for this entrepreneur is for them to self-finance the firm; or, since the entrepreneur here is not wealthy enough to completely self-finance, they should get as close to self-financing as possible by putting all of their money into the firm. There is nothing inherent in liability structures preventing the entrepreneur from doing this under limited liability. As 3.2.1 shows, synthetic unlimited liability is costless to create in a firm incorporating under limited liability.

Interpreting 3.2.2 in the 'opposite' direction shows another facet in the debate between forms of liability: that unlimited liability firms are also prone to these moral hazard issues. Although these moral hazard problems are often studied in the context of limited liability with outside wealth, they exist in unlimited liability firms as well. To some extent liability in the modern world is always 'limited' to draining the entrepreneur's wealth. Excess nonpecuniary punishments such as debtors' prison have long been outlawed in most nations. But even considering moral hazard in extreme cases where, for example, debtors' prison exists, it's clear that this extra punishment does not eliminate moral hazard but instead shifts the threshold around which the action is taken. If a firm is doing so poorly that without risk shifting its owner will go to prison with certainty, they will engage in this value destroying activity identically to a modern counterpart who considers the actions in reference to their level of protected outside wealth.

Once in place, limited liability can cause a number of moral hazard issues. But ex-ante, nothing about the liability structure is inherently linked with these types of moral hazard, and once in place, even unlimited liability firms have the same set of problems as well. The only difference is behavior changes on different thresholds between unlimited liability and limited liability with positive outwside wealth $W_i - A_i > 0$.

3.3 Illiquidity of wealth

A reasonable concern is that these results have so far relied on the assumption that wealth is costless to liquidate. Suppose liquidating an entrepreneur's initial wealth is costly; for every \$1 of wealth they liquidate they receive $\ell < 1$ dollars of cash they can pay towards the initial investment cost. This can be motivated by the simple observation that most people have the majority of their wealth in illiquid assets, such as homes and cars. These are costly to liquidate as well as costly in the nonpecuniary sense that entrepreneurs derive utility from using these assets.

An immediate consequence of introducing costly liquidation is that *contingent* compensation dominates *non-contingent* compensation. Surplus is maximized when expected liquidation costs are minimized, and expected liquidation costs are minimized when liquidation is state contingent. An entrepreneur liquidating their wealth at the beginning of the project to pay I will incur liquidation costs with certainty, while entrepreneurs liquidating their wealth to repay creditors only when the firm fails will incur liquidation costs with some probability less than one. Then, a risk neutral entrepreneur would prefer to choose $A_i^* = 0$. Creditors pay the entire initial investment I in some kind of unlimited liability-type payment structure. The entrepreneur is strictly better off if they only pay creditors when the firm fails (liquidation costs paid with probability $(1 - p_H)$ compared to ex-ante liquidation for $A_i^* > 0$ (liquidation costs paid with probability 1). In this setting, unlimited liability dominates limited liability for a risk neutral investor. Limited liability by definition protects an entrepreneur's outside wealth when the firm fails, requiring the entrepreneur to liquidate their wealth before the firm begins in order to achieve a high enough stake to satisfy moral hazard.

However, this definition of limited liability is fairly strict. Instead of limited liability protecting an entrepreneur's outside wealth when the firm fails, limited liability has become a restriction on promising to repay *anything* when the firm fails. A more flexible definition of limited liability which retains some amount of wealth protection but allows the entrepreneur to participate in contingent compensation is needed.

To solve the problem of costly liquidation in a limited liability framework, I will add a common provision in debt contracts - collateral pledging. When an asset is pledged as collateral to creditors, it can be used to repay the face value of debt when the firm fails even under the limited liability firm structure. Doing so would allow an entrepreneur to, for example, pledge their car as collateral but protect their debit account at a bank. In the context of the model, it allows them to contract contingent payments for any value equal to or less than $\ell(W_i - A_i)$ under limited liability. This more realistic definition of limited liability once again restores limited liability to the optimal contract.

Proposition 4 *With liquidation costs to wealth, possible repayments to creditors allowing the entrepreneur to satisfy moral hazard under limited liability with collateral pledging is strictly greater than the possible repayments to creditors under unlimited liability.*

Once an entrepreneur financing with debt under a limited liability regime is allowed to contract on state contingent repayments the result that limited liability is the more flexible form of incorporation once again holds. Limited liability can costlessly replicate unlimited liability consumption outcomes (by pledging an entrepreneur's entire outside wealth as collateral), but the opposite is not true: replicating limited liability consumption outcomes in an unlimited liability regime again relies on equity financing, which is costly for moral hazard.

3.4 Application to Barriers of Entry

The cost of starting a new limited liability firm varies considerably around the world. This can take place in either explicit registration costs or in implicit costs such as higher minimum sizes for the firm. As stated in the introduction, in the UK, a limited liability company can be set up for £12 (or £30 for expedited processing) and no minimum capital: any amount about £0 is sufficient. In contrast, Italy requires €10,000 in initial capital and around €2,500 in fees. Outside of Europe, some countries have even higher barriers to entry. In Mozambique, fees alone are estimated to be around $20,000 to begin a limited liability company.

Several papers attempt to empirically test whether having high barriers to entry for limited liability firms is in the public interest, e.g. Djankov et al. (2002). Suppose that, in order to keep low quality or fraudulent entrepreneurs from beginning firms, the limited liability firm type must be hard to access. That would suggest that countries with high barriers to entry would have fewer firms, but higher quality firms. Djankov et al. (2002) finds that higher barriers to entry are not associated with higher quality products, fewer damaging externalities, or greater competition. In contrast, they find high barriers to entry are associated with higher levels of corruption and greater size of the unofficial economy. Taken together, it appears that high barriers to entry are more about protecting rents than in protecting public interest.

In a similar vein, Klapper et al. (2006) studies the distribution of firm productivity and firm size in countries with different barriers of entry. They find that countries with high barriers of entry have i) lower creation of new firms, ii) higher size of new firms (consistent with the hypothesis that a firm would have to be very large to pay the entry fees) and iii) lower productivity growth of incumbent firms. This again suggests that barriers to entry are not to prevent low quality entrepreneurs from entering, but to protect incumbents from competition.

The results in prior sections of this paper can help identify the implications of high barriers to entry: a high barrier to entry is an unnecessary restriction on entrepreneur participation. By requiring entrepreneurs to pay a steep fee to access the largest possible choice of repayments to creditors, countries discourage entrepreneurship, not just for poor quality entrepreneurs who are supposedly targeted by these policies. It pushes all entrepreneurs towards the unlimited liability firm types, which mandate maximum punishment for the failure of their business. To the extent that entrepreneurs are risk averse, as has been documented in several other studies (see e.g. Gottlieb et al. (2022), Fan and White (2003)), unlimited liability is costly for entrepreneurs.

Making limited liability easier to access is only one way in which governments can choose to protect the wealth of entrepreneurs when their firm does not succeed. Another common policy tool is bankruptcy. With generous bankruptcy, access to limited liability is perhaps unnecessary. However, studies of bankruptcy find that, unlike low barriers to entry, generous bankruptcy has mixed impacts on entrepreneurship. Berkowitz and White (2002) find that states with more generous bankruptcy are also associated with entrepreneurial credit rationing; entrepreneurs are less likely to find a loan, and conditional on finding a loan, it will be at a higher interest rate.

To identify the implications of making bankruptcy more generous in the entrepreneur's choice of contracts, I will add bankruptcy protections to my analysis under liquid and illiquid wealth.

3.4.1 Bankruptcy

I represent bankruptcy as a threshold under which outside wealth is protected regardless of the entrepreneur's choice of liability. This threshold is denoted as b. An entrepreneur with outside wealth $W_i - A_i$ can declare bankruptcy when the firm fails and protect up to b of this wealth from creditors regardless of their choice of liability type or previous pledging for collateral. For simplicity, consider no cost to declaring bankruptcy.

Both access to limited liability and bankruptcy ease the entrepreneur's risk aversion by allowing them to retain some consumption when their firm fails. In this way they are often viewed as close substitutes as policy tools to cater to entrepreneur's risk aversion and encourage entrepreneur participation. To analyze bankruptcy, I will return to the simplest assumption, that wealth is fully liquid. In a similar fashion to earlier in the paper, I will then try to construct synthetic bankruptcy using limited liability and synthetic limited liability using bankruptcy and unlimited liability.

Proposition 5 *If entrepreneur wealth and choice of stake satisfy $W_i - A_i \leq b$, synthetic limited liability is costless.*

Proposition 6 *For any threshold b, creating synthetic bankruptcy with limited liability is costless.*

Corollary 2 *The choice of creditor repayments satisfying moral hazard under limited liability is strictly larger than the choice of repayments under bankruptcy and unlimited liability $\forall b < \infty$.*

With liquid wealth, bankruptcy and limited liability are indeed close substitutes. Both protect the outside wealth of the entrepreneur. In the first proposition, synthetic limited liability can be created by incorporating under unlimited liability, choosing debt financing, and declaring bankruptcy to protect all outside wealth any time that the firm's value is below F. However, for a sufficiently wealthy entrepreneur, or one that needs a very small stake to overcome moral hazard, the bankruptcy threshold is too low to be meaningful. It can only fully replicate limited liability when the threshold $b = \infty$. In contrast, synthetic bankruptcy with limited liability is always costless by simply leaving the amount of wealth the entrepreneur would like protected outside of the firm. Then, bankruptcy plus unlimited liability offer a smaller choice of stake for some entrepreneurs than limited liability with a finite bankruptcy threshold b.

Using the infinite bankruptcy threshold is still compatible with all choices of stake $A_i \leq W_i$: wealth liquidated ex-ante and added to the firm can still be lost and enters the IC constraint. The infinite threshold for bankruptcy protects the entrepreneur from being over-punished from being i) too wealthy or ii) having only a small moral hazard such that A^{IC}_{min} is low, making some choices of outside wealth $W_i - A^*_i > b$, $A^*_i \geq A^{IC}_{min}$ above the bankruptcy threshold and thus accessible for creditors under unlimited liability debt financing.

To the extent that limited liability regimes are made expensive by governments erecting high barriers to entry, making bankruptcy more generous encourages participation by making synthetic limited liability easier to access. This is exactly the empirical finding in Armour and Cumming (2008): there is a strong, positive interaction term between generous bankruptcy and high minimum capital requirements for limited liability firms when seeking to explain entrepreneur participation. That is, bankruptcy is more impactful for encouraging entrepreneur participation when minimum capital requirements for limited liability firm types is high. Cheap synthetic limited liability is desirable when limited liability is otherwise costly.

The previous propositions and corollary demonstrate the substitutability of limited liability and bankruptcy under liquid wealth. When illiquid wealth is introduced, bankruptcy becomes a restriction of contract choice rather than a potentially redundant way of accessing limited liability protections.

Proposition 7 *With illiquid wealth, bankruptcy strictly reduces the choice of contracts under both limited and unlimited liability.*

Unlike limited liability, which can have exceptions with pledged collateral, an entrepreneur cannot pledge assets in the event of bankruptcy. Under liquid wealth this restriction is not meaningful. However, with illiquid wealth, this is not an innocuous implication that eliminates redundant contracts in common with limited liability. By reducing the ability of the entrepreneur to contract contingent payments, it punishes an entrepreneur with low outside wealth. In other words, it forces entrepreneurs to liquidate and commit wealth ex-ante to the firm to achieve a high stake $A_i^* \geq A_{min}^{IC}$.

To demonstrate this proposition, consider an entrepreneur who wishes to pledge all of their illiquid assets worth $W_i = \$b$ as collateral to their firm's debt (that is, pay up to $\$\ell b$ when the firm produces insufficient cashflows). Without bankruptcy, choosing $A_i^* = 0$ could be done under both unlimited liability or under limited liability with collateral pledging. In either case the assets are only liquidated when the firm fails, providing access valuable contingent compensation when wealth is illiquid. When bankruptcy protects b of the entrepreneur's wealth, pledging the assets as collateral or having it as a personal asset to support debt repayment under unlimited liability is no longer credible. The only way for the entrepreneur to use their wealth in illiquid assets to impact their ability to overcome moral hazard is to liquidate the assets ex-ante, gaining $\ell b < b$ of liquid wealth. Note that since the deadweight loss is not state contingent, it does not help overcome any moral hazard that requires skin in the game; it does not punish the entrepreneur for failure or reward them for success.

Bankruptcy can still, on net, encourage entrepreneur participation. To the extent that synthetic limited liability is desirable for risk averse entrepreneurs, the benefits may outweigh any potential deadweight losses from liquidation or loss in choice of stake for wealthy entrepreneurs, especially when the limited liability firm type is costly. But unlike lowering barriers to entry, bankruptcy is not a purely positive effect on entrepreneur participation. Demonstrating this result empirically, in Koudijs and Salisbury (2020), bankruptcy-like protections to a wife's property[10] from a husband's investments had a mixed impact on entrepreneurial activities.

[10] The authors call this protection 'limited liability' because it gave downside protection to the wife's assets. However, the inability to pledge assets as collateral more resembles modern bankruptcy: "This meant that married women's property could not be made liable for a loan." Koudijs and Salisbury (2020).

When husbands were richer, it encouraged investment by giving access to cheap synthetic limited liability, while when husbands were poorer, it discouraged investment by decreasing the ability to contract state contingent repayments with illiquid wealth.

3.5 Other Considerations

3.5.1 The prevalence of unlimited liability

Finally, this paper has offered many reasons why limited liability seems to be the superior corporate form. So why is there such a large number of unlimited liability firms in reality? Typically, these make up the majority of registered small businesses in any state or country. The appeal of unlimited liability is its simplicity and the ease at which a firm is created. Unlimited liability firms are almost universally cheaper to create and have lower regulatory reporting requirements.

To the extent that unlimited liability is undesirable among many different shareholders (for discussion see e.g. Winton (1993), or Easterbrook and Fischel (1985)), entrepreneurial firms expecting to grow and go public will prefer limited liability while small firms owned by a single person may be indifferent between the kind of liability. Unlimited liability can be seen as having an implicit restriction in firm size, as after a certain point, limited liability equity financing with public markets becomes the best way to raise large amounts of capital.

The level of debt is also important. Entrepreneurs who take out no debt and require no financing will be indifferent between types of liability, and many will naturally choose the simpler form of liability. Somebody who wants to start a business providing services such as tutoring or editing will have very little to no startup costs and have no need to incur debt. To the extent that it is usually quicker and cheaper to set up unlimited liability firms, these entrepreneurs will choose unlimited liability.

Finally, as has been shown in the discussion of bankruptcy, entrepreneurs with little to no outside wealth may rely on bankruptcy protections rather than limited liability in order to protect their assets outside the firm. It may be less costly to declare bankruptcy when a firm fails than it is to pay to set up a limited liability firm.

3.5.2 Intermediate liability regimes

Historically, there have been regimes between full limited and full unlimited liability. For example, with double liability, a shareholder would be required to pay up to the book value of their share when the firm fails; a share worth $100 would require their shareholder to inject $100 of money into the failed firm. This was used primarily in banking to provide some guarantee to depositors as opposed to regular firms providing guarantees to creditors. Several papers have studied these intermediate regimes: see, for example, Grossman (2001), Bodenhorn (2015), or Anderson et al. (2018).

These regimes are redundant in my setting. With liquid wealth, the double liability obligation can be liquidated and added to the firm ex-ante. With illiquid wealth, the contract can be costlessly replicated with limited liability plus collateral pledging.

Finally, the regimes retain many problems of unlimited liability. Problems with multiple shareholders with more than strictly limited liability apply regardless of if the liability is a fixed amount or is unlimited: a shareholder with any extra liability may gift all their wealth to a nephew and 'be penniless' when it is time to pay their extra obligation. Intermediate liability was 'joint and several', just like unlimited liability, meaning shareholders may owe more than the book value of their share if other shareholders are too poor to fulfill their obligations. This incentivizes monitoring of other sharedholders' wealth identically to unlimited liability.

High liquidity of shares eventually undermined double liability; for a full discussion, see Macey and Miller (1992).

3.6 Conclusion

By examining an entrepreneur's choice of contracts under limited liability, unlimited liability, and bankruptcy, this paper provides an explanation for observed empirical trends around high barriers to entry for limited liability firms and around making bankruptcy policy more generous. With both liquid and illiquid wealth, limited liability allows entrepreneurs with sufficient wealth a strictly larger choice of creditor repayments which solve moral hazard. Synthetic unlimited liability is always costless, while the opposite is not true; synthetic limited liability requires equity financing, which is costly for moral hazard. Although limited liability protects outside wealth it is compatible with any level of punishment for firm failure, including the maximum mandated by unlimited liability.

Because synthetic unlimited liability is always costless, limited liability can solve moral hazard problems equally as well as unlimited liability. Any moral hazard problem requiring 'skin in the game' from the entrepreneur can be solved by both liability regimes, or neither. Then, it is unrealistic to think that countries that allow easy access to limited liability allow for poor quality entrepreneurs to flourish. By having high explicit or implicit costs of entry for limited liability firms, policymakers restrict access to the full choice of repayments for new entrepreneurs and push them towards costly unlimited liability firm types that mandate maximum punishment.

When forms of liability are equally costly and wealth is fully liquid, bankruptcy is redundant with limited liability. When limited liability is more expensive to access, generous bankruptcy becomes a cheaper tool for synthetic limited liability than equity financing, encouraging entrepreneur participation. The positive impact of making bankruptcy generous rises with high barriers to entry. When wealth is illiquid, bankruptcy represents a strict reduction in the entrepreneur's choice of contracts by eliminating state contingent liquidation and requiring costly ex-ante liquidation of wealth to satisfy skin in the game requirements. This reduction in the choice of contracts impacts both limited and unlimited liability firm types, and negatively impacts entrepreneurs' ability to negotiate for credit. However, if high barriers to entry are in place around limited liability firm types, bankruptcy can still have a positive net impact on entrepreneur participation.

3.7 Appendix: Proofs for Entrepreneur Participation with Limited Liability and Bankruptcy

3.7.1 Proof of proposition 1: synthetic limited liability

For the first part of the proposition, note that forms of liability only impact debt payments. A firm financed with equity has identical payouts to the creditor and entrepreneur under both limited and unlimited liability.

For the second part of the proposition, that financing under equity requires a higher stake than financing under debt, I will demonstrate that the current model does have a lower minimum stake under debt financing than equity financing. Taking the IC and break-even constraint for debt:

$$p_H(R-F)+(W_i-A_i)\geq p_L(R-F)+(W_I-A_i)+B \tag{44}$$

$$I-A=p_HF+(1-p_H)r \tag{45}$$

Plugging the break-even constraint into the IC constraint yields

$$\rightarrow A^{IC}_{min}(D)=p_H\frac{B}{\Delta p}-(p_HR-I)-(1-p_H)r \tag{46}$$

Similarly for equity,

$$(\alpha)p_HR+(\alpha)(1-p_H)r\geq(\alpha)p_LR+(\alpha)(1-p_L)r+B \tag{47}$$

$$(1-\alpha)p_HR+(1-\alpha)(1-p_H)r\geq I-A_i \tag{48}$$

$$\rightarrow A^{IC}_{min}(E)=[\frac{B}{\Delta p(R-r)}-1](r+p_H(R-r))+I \tag{49}$$

Comparing the two,

$$p_H\frac{B}{\Delta p}-(p_HR-I)-(1-p_H)r<[\frac{B}{\Delta p(R-r)}-1](r+p_H(R-r))+I \tag{50}$$

$$\rightarrow p_H\frac{B}{(\Delta p(r+p_H(R-r))}<p_H\frac{B}{\Delta p(p_H(R-r))} \tag{51}$$

Which holds as $r>0$.

3.7.2 Proof of proposition 2: synthetic unlimited liability

First, note that the entrepreneur's choice of stake is irrelevant under unlimited liability. Since $W_i<\frac{I-r}{1+p_H}$, the entrepreneur cannot finance debt with $F<r+$

W_i. Then, even if they choose $A_i = 0$ and keep all of their wealth outside the firm, they will still consume zero if the firm fails. Since the consumption of the entrepreneur is zero when the firm fails and creditors are risk neutral and competitive the entrepreneur must consume $\frac{p_H R+(1-p_H)r-I}{p_H} + W_i$ when the firm succeeds: the entrepreneur claims the entire firm NPV in expectation under any contract, and since their low state consumption is zero, their high state consumption must contain the entire expected NPV plus their initial wealth. This high state repayment is feasible because the firm has positive NPV when the IC constraint is satisfied. Since the entrepreneur consumes the same under both states for any choice of $A_i \leq W_i$ the face value of debt is unique and fixed at

$$F = \frac{I - W_i - (1-p_H)r}{p_H} \tag{52}$$

To summarize, because the face value of the debt contract and all consumption outcomes are identical for any choice of $A^{IC}_{min}(D) \leq A^*_i \leq W_i$ under unlimited liability any choice of stake under unlimited liability can be represented with the choice $A^*_i = W_i$.

Finally, note that when $A^*_i = W_i$ the entrepreneur has no outside wealth and therefore liability regimes are equivalent. Repayment from the entrepreneur's personal wealth to creditors when the firm fails under either regime is zero when $A^*_i = W_i$. Then, a firm can incorporate under limited liability and choose $A^*_i = W_i$ to costlessly replicated the unlimited liability debt contract and consumption outcomes, which is the definition of synthetic unlimited liability.

3.7.3 Proof of corollary 1: limited liability admits larger choice of repayments than unlimited liability

With liquid wealth, synthetic unlimited liability is costless: a firm incorporating under limited liability simply chooses $A^*_i = W_i$. However, limited liability is compatible with any $A^{IC}_{min}(D) < A^*_i \leq W_i$. Since $A^{IC}_{min}(D) < W_i$ (Assumption 3), limited liability admits the strictly larger choice of A^*_i i.e. the larger set of contracts.

3.7.4 Proof of corollary 1: moral hazard

If a moral hazard problem can be solved by $A^*_i \geq A^{IC}_{min}$, it can be solved equally well by unlimited liability (where $A^*_i = W_i$) or by limited liability by choosing $A^*_i = W_i$.

3.7.5 Proof of proposition 4: illiquid wealth

Unlike the previous case with liquid wealth, the entrepreneur no longer has identical consumption between choosing any $A^*_i \leq W_i$ with unlimited liability. For any $A^*_i > 0$ the entrepreneur must liquidate $\frac{A^*_i}{\ell}$ ex-ante, incurring liquidation costs with probability 1. For $A^*_i = 0$ the entrepreneur only incurs deadweight liquidation costs

$(1-\ell)W_i$ when the firm fails with probability $1-p_H$. A risk neutral creditor acting optimally would choose $A_i^* = 0$. However, if an entrepreneur with a different utility function were to choose $A_i^* > 0$ they would owe the creditor $r + \ell(W_i - A_i)$ when the firm fails, since $F > r + W_i$ (Assumption 2). In other words, regardless of their choice of A_i^*, with unlimited liability, the entrepreneur consumes zero in the low state.

When collateral pledging is introduced to limited liability it allows the entrepreneur to voluntarily contract around their limited liability: ex-ante they are allowed to choose any collateral payment $C \leq \ell(W_i - A_i^*)$ to the creditors when the firm fails. Identically to the unlimited liability problem a risk neutral entrepreneur would optimally choose $A_i^* = 0$. However, an entrepreneur with a different utility function operating under limited liability and collateral pledging is able to choose to have consumption greater than zero in the low state. They would do so by choosing $A_i^* \leq \ell W_i$ and collateral repayment $C < x(W_i - A_i^*)$ subject to incentive compatibility and break-even constraints.

To summarize, an entrepreneur financing with debt under unlimited liability can never contract to have positive consumption when the firm fails. An entrepreneur with limited liability and collateral pledging can contract for consumption when the firm fails. The limited liability firm can additionally replicate any unlimited liability contract by choosing $C = x(W_i - A_i^*)$. Limited liability once again admits the larger choice of repayments.

3.7.6 Proof of proposition 5: synthetic LL with bankruptcy

Any entrepreneur incorporating under unlimited liability with $W_i - A_i^* < b$ will pay creditors zero from their personal wealth in the low state (when the firm fails): all outside wealth is protected by bankruptcy law. This is equivalent to financing with limited liability, when any $W_i - A_i^*$ also results in the creditors only receiving r in the low state. If the low state repayment of creditors is identical the high state repayment of creditors (when the firm succeeds) is identical through the break even constraint. If both repayments to creditors are the same the consumption of entrepreneurs is identical under both unlimited liability with bankruptcy protections and limited liability.

3.7.7 Proof of proposition 6: synthetic bankruptcy

For any finite b, an entrepreneur choosing to incorporate under unlimited liability with $W_i - A_i^* < b$ can instead choose to incorporate under limited liability. In both cases low state repayment to creditors from the entrepreneur's personal wealth is zero, as all outside wealth is protected by definition under both regimes. If the low state repayment of creditors is identical the high state repayment of creditors is identical. If both repayments are identical the consumption of entrepreneurs is identical under both limited liability or unlimited liability with sufficiently high bankruptcy protections.

3.7.8 Proof of corollary 2: limited liability has a larger choice of creditor repayments than unlimited liability with bankruptcy

For any finite b an entrepreneur with $W_i - A_i^* > b$ cannot protect their wealth from creditors. The creditors would receive in total $(W_i - A_i^*) - b + r$ in the low state. An entrepreneur with limited liability and no bankruptcy protections can costlessly replicate this by choosing collateral $C = (W_i - A_i^*) - b$, and can additionally choose $C = 0$, meaning creditors are only repaid r from firm revenues in the low state. A finite bankruptcy threshold b by definition fails to protect any outside wealth above b, which is always possible for limited liability: limited liability admits the larger choice of repayments for sufficiently wealthy entrepreneurs.

3.7.9 Proof of proposition 7: bankruptcy reducing the choice of repayments with illiquid wealth

The legal representation of bankruptcy in the model is that the bankruptcy threshold b cannot be contracted around, even by explicit collateral pledging. For example, if bankruptcy protects an entrepreneur's house, they cannot credibly offer the house as collateral to repay a loan.

With this representation bankruptcy is a restriction on state contingent repayments. An entrepreneur with unlimited liability cannot choose $A_i^* = 0$ to repay creditors xW_i when the firm fails. They have to choose state contingent repayment $C \leq \ell((W_i - A_i^*) - b)$ Likewise, an entrepreneur with limited liability cannot pledge their entire wealth as collateral: they can no longer choose $A_i^* = 0, C = \ell W_i$. Since bankruptcy places restrictions on state contingent repayments, and state contingent repayments are strictly preferable to ex-ante liquidation, bankruptcy reduces the choice of repayments.

References

Admati, Anat R and Stephen A Ross, "Measuring investment performance in a rational expectations equilibrium model," *Journal of Business*, 1985, pp. 1–26.

Agarwal, Vikas, Gerald D Gay, and Leng Ling, "Window dressing in mutual funds," *The Review of Financial Studies*, 2014, *27* (11), 3133–3170.

Anderson, Haelim, Daniel Barth, and Dong Beom Choi, "Reducing moral hazard at the expense of market discipline: the effectiveness of double liability before and during the Great Depression," *Office of Financial Research Research Paper*, 2018, (18-06).

Ang, Andrew, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang, "The cross-section of volatility and expected returns," *The Journal of Finance*, 2006, *61* (1), 259–299.

Armour, John and Douglas Cumming, "Bankruptcy law and entrepreneurship," *American Law and Economics Review*, 2008, *10* (2), 303–350.

Barber, Brad M, Xing Huang, and Terrance Odean, "Which factors matter to investors? Evidence from mutual fund flows," *The Review of Financial Studies*, 2016, *29* (10), 2600–2642.

Barras, Laurent, Olivier Scaillet, and Russ Wermers, "False discoveries in mutual fund performance: Measuring luck in estimated alphas," *The journal of finance*, 2010, *65* (1), 179–216.

Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song, "What do mutual fund investors really care about?," *The Review of Financial Studies*, 2022, *35* (4), 1723–1774.

Berk, Jonathan B and Jules H Van Binsbergen, "Measuring skill in the mutual fund industry," *Journal of financial economics*, 2015, *118* (1), 1–20.

_ **and Richard C Green**, "Mutual fund flows and performance in rational markets," *Journal of political economy*, 2004, *112* (6), 1269–1295.

Berkowitz, Jeremy and Michelle J White, "Bankruptcy and small firms' access to credit," 2002.

Bhattacharya, Sudipto, Amil Dasgupta, Alexander Guembel, and Andrea Prat, "Incentives in Funds Management: A Literature Review," *Handbook of Financial Intermediation and Banking, ed. A*, 2008.

Bodenhorn, Howard, "Double liability at early American banks," Technical Report, National Bureau of Economic Research 2015.

Brown, Keith C, W Van Harlow, and Laura T Starks, "Of tournaments and temptations: An analysis of managerial incentives in the mutual fund industry," *The Journal of Finance*, 1996, *51* (1), 85–110.

Carhart, Mark M, "On persistence in mutual fund performance," *The Journal of Finance*, 1997, *52* (1), 57–82.

Chernenko, Sergey and Adi Sunderam, "Liquidity transformation in asset management: Evidence from the cash holdings of mutual funds," Technical Report, National Bureau of Economic Research 2016.

Chevalier, Judith and Glenn Ellison, "Risk taking by mutual funds as a response to incentives," *Journal of political economy*, 1997, *105* (6), 1167–1200.

Cleveland, William S and Susan J Devlin, "Locally weighted regression: an approach to regression analysis by local fitting," *Journal of the American statistical association*, 1988, *83* (403), 596–610.

Cuthbertson, Keith, Dirk Nitzsche, and Niall O'Sullivan, "Mutual fund performance: Measurement and Evidence 1," *Financial Markets, Institutions & Instruments*, 2010, *19* (2), 95–187.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, "Measuring mutual fund performance with characteristic-based benchmarks," *The Journal of finance*, 1997, *52* (3), 1035–1058.

Dasgupta, Amil and Andrea Prat, "Financial equilibrium with career concerns," *Theoretical Economics*, 2006, *1* (1), 67–93.

_ and _ , "Information aggregation in financial markets with career concerns," *Journal of Economic Theory*, 2008, *143* (1), 83–113.

_ and Giorgia Piacentino, "The wall street walk when blockholders compete for flows," *The Journal of Finance*, 2015, *70* (6), 2853–2896.

_ , Vyacheslav Fos, Zacharias Sautner et al., "Institutional investors and corporate governance," *Foundations and Trends in Finance*, 2021, *12* (4), 276–394.

Djankov, Simeon, Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer, "The regulation of entry," *The quarterly Journal of economics*, 2002, *117* (1), 1–37.

Easterbrook, Frank H and Daniel R Fischel, "Limited liability and the corporation," *U. Chi. L. Rev.*, 1985, *52*, 89.

Elton, Edwin J, Martin J Gruber, and Christopher R Blake, "A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and Morningstar mutual fund databases," *The Journal of Finance*, 2001, *56* (6), 2415–2430.

Fama, Eugene F and Kenneth R French, "A five-factor asset pricing model," *Journal of Financial Economics*, 2015, *116* (1), 1–22.

Fan, Wei and Michelle J White, "Personal bankruptcy and the level of entrepreneurial activity," *The Journal of Law and Economics*, 2003, *46* (2), 543–567.

French, Kenneth R., "Data Library," `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html` 2025. Accessed: 2025-01.

Gottlieb, Joshua D, Richard R Townsend, and Ting Xu, "Does career risk deter potential entrepreneurs?," *The Review of Financial Studies*, 2022, *35* (9), 3973–4015.

Gropp, Reint, John Karl Scholz, and Michelle J White, "Personal bankruptcy and credit supply and demand," *The Quarterly Journal of Economics*, 1997, *112* (1), 217–251.

Grossman, Richard S, "Double liability and bank risk taking," *Journal of money, credit and banking*, 2001, pp. 143–159.

Guerrieri, Veronica and Péter Kondor, "Fund managers, career concerns, and asset price volatility," *American Economic Review*, 2012, *102* (5), 1986–2017.

Halpern, Paul, Michael Trebilcock, and Stuart Turnbull, "An economic analysis of limited liability in corporation law," *U. Toronto LJ*, 1980, *30*, 117.

Holmstrom, Bengt and Jean Tirole, "Financial intermediation, loanable funds, and the real sector," *the Quarterly Journal of economics*, 1997, *112* (3), 663–691.

Jegadeesh, Narasimhan, "Evidence of predictable behavior of security returns," *The Journal of finance*, 1990, *45* (3), 881–898.

_ **and Sheridan Titman**, "Profitability of momentum strategies: An evaluation of alternative explanations," *The Journal of finance*, 2001, *56* (2), 699–720.

Jensen, Michael C and William H Meckling, "Theory of the firm: Managerial behavior, agency costs and ownership structure," in "Corporate governance," Gower, 1976, pp. 77–132.

Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng, "Unobserved actions of mutual funds," *The Review of Financial Studies*, 2008, *21* (6), 2379–2416.

Klapper, Leora, Luc Laeven, and Raghuram Rajan, "Entry regulation as a barrier to entrepreneurship," *Journal of financial economics*, 2006, *82* (3), 591–629.

Kosowski, Robert, "Do mutual funds perform when it matters most to investors? US mutual fund performance and risk in recessions and expansions," *The Quarterly Journal of Finance*, 2011, *1* (03), 607–664.

Koudijs, Peter and Laura Salisbury, "Limited liability and investment: Evidence from changes in marital property laws in the US South, 1840–1850," *Journal of Financial Economics*, 2020, *138* (1), 1–26.

Kyle, Albert S, "Continuous auctions and insider trading," *Econometrica: Journal of the Econometric Society*, 1985, pp. 1315–1335.

Lehmann, Bruce and Allan Timmermann, "Performance measurement and evaluation," 2007.

Lo, Andrew W, "The statistics of Sharpe ratios," *Financial analysts journal*, 2002, *58* (4), 36–52.

Lou, Dong, "A flow-based explanation for return predictability," *The Review of Financial Studies*, 2012, *25* (12), 3457–3489.

Lynch, Anthony W and David K Musto, "How investors interpret past fund returns," *The Journal of Finance*, 2003, *58* (5), 2033–2058.

Macey, Jonathan R and Geoffrey P Miller, "Double liability of bank shareholders: history and implications," *Wake Forest L. Rev.*, 1992, *27*, 31.

Rhoades, Stephen A, "The herfindahl-hirschman index," *Fed. Res. Bull.*, 1993, *79*, 188.

Thompson, Samuel B, "Simple formulas for standard errors that cluster by both firm and time," *Journal of Financial Economics*, 2011, *99* (1), 1–10.

Wermers, Russ, "A matter of style: the causes and consequences of style drift in institutional portfolios," Technical Report, CFR Working Paper 2012.

Winton, Andrew, "Limitation of Liability and the Ownership Structure of the Firm," *The Journal of Finance*, 1993, *48* (2), 487–512.