THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Essays in Financial Economics

ALBERTO PELLICIOLI

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

April 2021

To my parents, for their unconditional love and support.

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).

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

Statement of conjoint work I confirm that chapter 1 is jointly co-authored with Christian Julliard. I contributed 50% of the work for chapter 1.

I declare that my thesis consists of 20,387 words.

Acknowledgment

My most sincere gratitude to my supervisor Christian Julliard for his help, support and guidance during the Ph.D. His knowledge and research have been an invaluable source of inspiration, without which this thesis would not have been possible. Co-authoring a paper with Christian have taught me more than any academic project. I also thank Michela Verardo and Cameron Peng for all the enlightening discussions and the inspiring comments.

I would like to thank the faculty members at LSE, including Georgy Chabakauri, Thummim Cho, Dong Lou, Ian Martin, Martin Oehmke and Christopher Polk. I benefited from their classes, cutting-edge research and comments during seminars. I am also grateful to the administrative staff, particularly Mary Comben, for their help. I would also like to acknowledge the generous financial support from the ESRC and the LSE Finance Department.

My experience at LSE would not have been the same without my colleagues. I thank Agnese Carella, Lorenzo Bretscher, Jesus Gorrin, James Guo, Brandon Han, Zhongchen Hu, Muneaki Iwadate, Lukas Kremens, Francesco Nicolai, Dimitris Papadimitriou, Michael Punz, Bernardo Ricca, Simona Risteska, Petar Sabtchevsky, Una Savic, Karamfil Todorov, Xiang Yin and Yue Yuan for their help and friendship.

A special thanks to my friends, colleagues and flatmates Marco Pelosi and Fabrizio Core. I would not have made it without you, and I am forever indebted. To Jianxu, for your encouragement and patience throughout this journey. To Agnese, Guido, Alice, Iacopo, Marco, Francesca, Nunzia, Filippo, Lucio, Eva, Luca, Simone and Gianpaolo for your friendship.

Most importantly, I want to wholeheartedly thank my parents and my whole family, who have always been there for me. Without your love and support, I would have never become the person I am today.

Abstract

In the first chapter, co-authored with Dr. Christian Julliard, we study the impact of option expiration on underlying stock volatility. We find a negative direct effect on stock realized volatility and a positive and significant effect on stock implied volatility. Moreover, a positive spillover effect on stocks with no options expiring on a given expiration date is observed. Two possible explanations are discussed, namely investors' delta hedging and stock pinning around option expiration dates. Both seem to affect stock volatility. Finally, we implement a trading strategy that takes advantage of these findings.

The second chapter studies the investment behaviour of mutual funds during financial bubbles. I find that mutual funds over-invest in bubble sectors during the run-up and withdraw money right before the collapse. This result is robust across different benchmark specifications and across fund styles. I also document that this strategy generates a positive and significant alpha (4% on an annual basis), with respect to both a risk-neutral expectation and the Fama-French factors. The paper provides evidence supporting the theory that mutual funds ride the bubble rather than causing it. It also demonstrates that mutual fund holdings can predict the future returns of a sector over a short to medium horizon.

Building on the previous findings, the third chapter studies the fee setting behaviour of mutual funds during financial bubbles. It shows that, besides the well-documented persistence of fees, mutual funds charge higher fees during price run-ups. Two theoretical models support the finding: the first shows that investors' sensitivity to fees decreases during bubble episodes, while the second demonstrates that the increase in fees translates into a higher mark-up over marginal cost.

Contents

1 Stock Volatility on Option Expiration Dates **9**
1.1 Introduction . . . 9
1.2 Data . . . 14
1.3 Empirical evidence . . . 15
1.4 Possible Explanations . . . 20
1.4.1 Delta Hedging . . . 21
1.4.2 Pinning . . . 22
1.5 Trading Strategy . . . 22
1.6 Robustness checks . . . 25
1.7 Conclusion . . . 27

2 Mutual Funds' Behaviour during Financial Bubbles **29**
2.1 Introduction . . . 29
2.2 Data . . . 33
2.3 Bubble episodes . . . 33
2.4 Mutual fund behaviour . . . 35
2.5 Predictability . . . 39
2.6 Endogeneity . . . 40
2.7 Conclusion . . . 43

3 On the Time-Series Persistence of Mutual Fund Fees **45**
3.1 Introduction . . . 45
3.2 Data . . . 48
3.3 Fees . . . 49
3.4 A Model of Investors' Sensitivity to Fees . . . 52
3.4.1 Demand . . . 52
3.4.2 Supply . . . 53
3.4.3 Estimation . . . 54
3.5 Marginal Cost vs. Mark-Up . . . 57
3.5.1 Marginal Cost . . . 57
3.5.2 Demand function . . . 58
3.5.3 Utility maximization . . . 58

3.5.4 Estimation 61
3.6 Conclusion 63

A Appendix to Stock Volatility on Option Expiration Dates 71

B Appendix to Mutual Funds' Behaviour during Financial Bubbles 84

C Appendix to On the Time-Series Persistence of Mutual Fund Fees 95

List of Tables

A1 Correlation among Volatility Measures . . . 71
A2 Regression of Volatility on Option Expiration Dates . . . 71
A3 Spillover Effect . . . 73
A4 Regression of Volatility on Open-Interest-weighted Delta Hedging 74
A5 Trading Strategy . . . 75
A6 Trading Strategy - Robustness Checks . . . 76
A7 Robustness Check: controlling for VIX . . . 77
A8 Robustness Check: controlling for Net Buying Pressure . . . 78

B1 Bubble Episodes . . . 85
B2 Excess Holdings . . . 89
B3 Stock Picking Ability (holdings) . . . 90
B4 Stock Picking Ability (trades) . . . 91
B5 Mutual Fund Excess Returns . . . 92
B6 Excess Holdings (value-weighted) . . . 93
B7 Contemporaneous Correlation between Excess Returns and Trades 93
B8 Return Predictability of Mutual Fund Holdings . . . 94

C1 Sector Classification . . . 96
C2 Summary Statistics . . . 97
C3 Fees during Bubble Episodes . . . 97
C4 Fees during Bubble Episodes (controlling for competition) . . . 98
C5 Regression of Market Share on Fees . . . 98
C6 IV Regression of Change in Market Share . . . 99
C7 IV Regression of Market Share . . . 100
C8 GMM Estimation . . . 101

List of Figures

A1 Volatility as a Function of Moneyness 72
A2 Implied Volatility around Expiration Dates 79
A3 Boxplot of Distance of Stock Price from Strike around Expiration 80
A4 Payoff Structure of a Straddle 81
A5 Volatility Smile of Call Options (controlling for VIX) 82
A6 Volatility Smile of Put Options (controlling for VIX) 83

B1 Timing of a Bubble Episode 84
B2 Evolution of Excess Holdings during a Bubble Episode 86
B3 Evolution of Excess Holdings during a Bubble Episode 87
B4 Evolution of Excess Holdings during a Bubble Episode. 88

1. Stock Volatility on Option Expiration Dates

1.1 Introduction

Understanding the interactions between equity and derivative markets is of paramount importance for financial research. Not only has the derivative market been growing continuously over the past 25 years[1], but also a shock in the option market can easily spills over into the underlying stock market.

Delta hedging (or, equivalently, option replication) is surely the most direct channel linking the stock and the option markets. Consider, for example, an investor who has a short position in a call option. She may want to hedge her exposure by buying the underlying asset, generating a buying pressure in the underlying stock market. Similarly, instead of buying a put option, a hypothetical investor may decide to sell the underlying asset and invest the proceeds at the risk-free rate in order to match the delta of the put option she would like to purchase. Historically, we know that option replication can have a significant impact on the stock market. In fact, in the late 1980s investors used to hedge against market risk by selling futures on the market index or, alternatively, by constructing synthetic put options by short-selling stocks. However, this was a pro-cyclical strategy, which forced investors to sell stocks when their price dropped, which in turn pushed down the price even more. On the 19th October 1987, the equity market experienced the largest negative return in its history (-22%), during the so-called Black Monday.

One could argue that delta hedging should not have an impact on the stock market, simply because long and short positions should offset each other. Hence, options may have an effect on the trading volume, but not on the first and second moments of the return distribution. This is true in complete and efficient markets only, where options are in zero net supply and agents hedge their portfolios. Clearly, financial markets do not meet these requirements: as noted by Lakonishok et al. (2006), purchased option positions are

[1] The Bank of International Settlement estimates the outstanding notional amount of the option market to be ten times that of the world stock market

less frequent than written positions for nonmarket maker investors; as a consequence, market makers have more purchased than written open interest. Additionally, non-market makers have four times more purchased calls than puts. Furthermore, while institutional investors are more prone to hedging strategies, speculators prefer not to. As a result, a well-documented option imbalance translates easily into an imbalance in the underlying stock market (Hu (2014)).

On option expiration dates, the interactions between option and stock markets are even stronger, due to portfolio managers trading activities. In fact, on top of dismissing the hedge on expired options and the change in option imbalance due to an option supply shock, portfolio managers may need to actively trade either to rebalance their portfolios and/or for liquidity needs. Indeed, equity options have *physical delivery*, meaning that there must be an exchange of assets if the option is exercised. Consider, for example, a portfolio manager with a target stock/bond ratio, who also has in-the-money call options on a given expiration date. Exercising the options would result in a purchase (at a discount) of stocks, which affects the portfolio composition. To restore the targeted stock/bond combination, she needs to sell the recently acquired stocks in the market. Similarly, if the same portfolio managers does not have enough liquidity in her portfolio to exercise the options, she may need to borrow and immediately re-sell the acquired stocks to pay off her loan. In both cases, option expiration triggers additional trading in the underlying stock market.

Due to continuous trading, it is extremely hard to estimate the effect of option on stock market throughout the option's life, as there are many other factors affecting the stock price whose impact cannot be accounted for easily. For this reason, the financial literature has focused either on option listing and/or option expiration dates. Option listing dates are particularly indicated for event studies as they represent a good threshold between before-the-event period and after-the-event period. However, option listing dates are announced in advance, so any impact may be biased due to anticipation effects. Furthermore, for a stock to be optionable, several regulatory requirements need to be met; hence the pool of stocks listed on the option market could be significantly different from the set of stocks which are not. Any event study based on option announcement or option issuance may suffer from selection bias. Conversely, expiration dates, which occur on the Saturday after the third Friday of each month, are defined for optionable stocks only, which constitute a homogeneous set of assets. Hence, no concern about a potential selection bias can be raised. In addition, stocks are still optionable after an expiration date. As there is no fundamental change in the asset, there shouldn't be any anticipation effect. For these reasons, expiration dates can be used to identify and estimate a relation between stock and option markets, which is the main goal of this paper. Besides, the impact of options on the underlying stock market is not only confined to the expected return but also to the volatility of returns. Although the literature is quite extensive, to the best of our knowledge there is no paper that studies the effect of option expiration dates on the variance of the stock return distribution. In fact, only Skinner (1989) shows that underlying stocks' volatility drops when options are listed; while what happens on option expiration dates is still uncertain. This paper fills this gap in the literature, focusing on four different volatility measures, two empirical and two implied from option prices, and it shows that, across all optionable stocks, implied volatility increases on option expiration dates, while empirical volatility drops. This hitherto unknown, although puzzling, phenomenon is paramount both for all market participants: for example, speculators may be able to gain for it, hedgers may need to adjust their hedge, regulators may want to be sure that no price manipulation occurs around expiration. Additionally, if optionable stocks are split between those who have an option expiring at a given expiration date and those who have not, we find that the latter group also experiences an increase in implied volatility; we refer to this as the *spillover effect*. We argue that these findings can be partially explained by investors' delta hedging and, potentially, by stock pinning. Pinning occurs when stock price tends to move closer to the option strike price around expiration. In this paper we provide evidence of pinning but we leave the discussion on its impact on stock volatility for future research. Moreover, we show that there is a strong relation between moneyness and volatility, which resembles the well-known "volatility smile" and persists after controlling for the overall market volatility as measured by the VIX index. Finally, to show that the findings discussed above are not only statistically but also economically significant, we construct a trading strategy that takes advantage of the discrepancy between realized and implied volatilities around expiration. This strategy generates an annualized Sharpe Ratio of 2.55.

Related Literature. Several authors attempted to estimate the relation between options and the underlying stocks. Recently, Hu (2014) shows that option trading generates imbalances in the stock market, mainly due to the delta hedging of the option market makers who operate in a less liquid market and want to reduce the risk of their short positions. He splits the stock imbalance in option-induced and non-option-induced stock imbalance, showing that the former positively predicts future returns. Besides this paper, the financial literature linking stock and option markets can be divided into four categories, depending on time (option listing vs expiration date) and the moment of the return distribution (return vs variance).

Although widely discussed in the financial literature, the effect of option issuance on the underlying stock return is still ambiguous. For example, Conrad (1989) shows that the introduction of options generates a permanent increase in price, beginning three days prior to listing. He argues that this is due to market makers making up their own inventory in anticipation of their need to delta hedge the short positions in the options. A broader effect on financial markets is found in Detemple & Jorion (1990), which shows that, around announcement, option listing produces an increase in the underlying stock market value as well as in the value of an industry index which excludes the recently optionable stocks. In other words, a spillover effect to other stocks within the same industry is observed when a stock becomes optionable. However, subsequent research shows that these results are not robust. Sorescu (2000) finds a negative price impact on stocks which become optionable after 1980, while Mayhew & Mihov (2005) finds that the effect of option listing on the underlying stock price vanishes when compared to a sample of control firms.

Regarding the second moment of the return distribution, when new options are introduced into the market, underlying stock volatility tends to decline in the order of 10-20% (Skinner (1989)), although the reason for this is still uncertain. French & Roll (1986) addresses the question of why volatility changes over time, especially during trading hours, identifying two possible reasons: a) stocks' returns are more volatile during trading hours due to a higher rate of information disclosure, and b) trading itself induces noise in stock price. However, Skinner (1989) finds no evidence of either of the two effects being related to option listing. Similarly, Bansal et al. (1989) shows that option listing leads to a decrease in the total risk of optionable stocks and to an increase in total trading volume. Conversely, Klemkowsky & Maness (1980), Trennepohl & Dukes (1979) and Whiteside et al. (1983) find little evidence of stock becom-

ing riskier after option listing. Finally, there is no strong evidence of option delisting affecting underlying stocks (Bartunek (1996)).
The second relevant day in an option's life is the expiration date, that is when investors close their positions by buying or selling the underlying asset. Although listing announcements may impact the stock market, it is actually at maturity that investors exchange assets, hence it is at maturity that the stock market should be impacted the most. In fact, Chiang (2014) finds that underlying stocks experience a negative return of 0.8% on option expiration dates, followed by slow reversal in the subsequent week. The author argues that this is due to the selling pressure of call holders who immediately sell the stock to either cash in their capital gains or for portfolio rebalancing reasons. This effect is stronger for stocks with higher open interest and more (deeply) in-the-money options. Also, there is no offsetting pressure from put holders, as the effect is even stronger when the net open interest (calls open interest minus puts open interest) is considered, nor when option writers are accounted for as their positions are usually hedged well before expiration. Similarly, Klemkosky (1978) finds an anticipation effect, with stock experiencing abnormal returns of -1% in the week before expiration and +0.4% in the week after.
This paper also relates closely to Ni et al. (2004), which shows that underlying stock prices converge to the options' nearest strike price around expiration and provide evidence of price manipulation by firm proprietary traders. The authors notice that the stock return is small in absolute value from the Thursday before the expiring Friday to the expiring Friday itself. Hence, stock prices stay close to the strike from the previous Thursday rather than converging to it only on the Friday. Differently from Chiang (2014), this phenomenon, called stock pinning[2], takes place in stocks with many slightly-in-the-money and slightly-out-of-the-money options. Interestingly, stock pinning is a much broader phenomenon that extends beyond the stock market. For example, Golez & Jackwerth (2012) provides evidence of pinning in the future market as well; in this case, the driver is not the time decay of delta hedge of market makers (in fact, they are actually short on index options) nor price manipulation (which is harder to implement in the future market), but investors having an incentive to sell their in-the-money call options or to early exercise them in order to avoid any price risk over the weekend. Finally, Chiang (2017) also

[2]Avellaneda et al. (2012) provide a theoretical framework that explains stock pinning on option expiration dates.

shows that trading volume spikes on option expiration dates. This paper also relates to the huge literature of stock volatility estimation, like Engle (1982), Koopman et al. (2005) and Schwert (1989) among the others. Finally, this paper relates also to Christensen & Prabhala (1998) which shows that implied volatility outperforms realized volatility as it both subsumes past information and forecasts future volatility. In this paper, we use both realized and implied volatility measures and we show that they behave differently around option expiration.

The chapter is organised as follows: section 1.2 describes the data used for the analysis and section 1.3 presents the main empirical results; possible explanations are discussed in section 1.4. Section 1.5 describes the trading strategy that exploits the main findings of the paper while robustness checks are reported in section 1.6. Section 1.7 concludes.

1.2 Data

Option data come from the IvyDB US database by OptionMetrics LLC. It contains data on all the options traded in the US market and reports historical price, expiration date, open interest, implied volatility and options type. The sample used in this paper covers the period from 1 January 1996 to 31 December 2006 in order not to overlap with the financial crisis of 2007 which had a significant impact on the overall market volatility.
Stock data come from the Center of Research in Security Prices (CRSP)[3] and include closing price, holding return, bid-ask spread and stock trading volume on a daily basis of all the stocks traded on the NYSE and AMEX for the same period. In order to eliminate any noise, penny stocks (those whose price is less than $5) are excluded from the sample. Furthermore, only stocks with at least 500 observations are included. For each stock, the intraday trades and quotes are obtained from TAQ. The CRSP database is augmented with daily data of the VIX index, a measure of overall market volatility from the Chicago Board Options Exchange. Due to the terrorist attack on September 11, 2001, no data is available for VIX on that day. However, options were traded in the market. After merging options and stocks data, we obtain a database with 903 option-

[3]Calculated (or Derived) based on data from database name ©2018 Center for Research in Security Prices (CRSP®), The University of Chicago Booth School of Business.

able stocks and 18,373,141 option trades, 11,170,815 of which are call option trades and the remaining 7,202,326 are put option trades. Call options account for 60.8% of the overall trading activity in the option market, while put options represent the remaining 39.2%. Finally, prices of futures on the S&P500 are obtained from Bloomberg while the historical list of S&P500 constituents is obtained from Compustat.

Options usually expire on the third Saturday of each month. Given that Saturdays are non-trading days and all the transactions must occur by the closing time of the previous day, we treat the third Friday as the expiration date. In the time frame considered, there are 132 expiration dates, among which 130 are on Fridays. The remaining two fell on Good Fridays (namely April 21, 2000 and April 18, 2003) and have been moved to the Thursday before. Moreover, a few options expired on the 29 September 2006 and on the 29 December 2006, but these options are excluded from the sample as they are not written on common stocks.

1.3 Empirical evidence

Stock volatility is defined as the fluctuations of returns around their mean. To address the impact of option expiration on the second moment of the underlying stock return distribution, four different volatility measures are constructed:

- *Squared returns:* starting for the formula of the variance

$$var(r_t) = \mathbb{E}[r_t^2] - \mathbb{E}[r_t]^2$$

  and using the fact that the daily mean stock return is usually very close to 0, the variance can be approximated by squared returns. The main limitation of this approach is that is doesn't account for past information. Volatility clustering[4] suggests that past volatility may be a good predictor of today's volatility and we may want to include it in our estimation.

- *GARCH:* for each stock a GARCH(1,1) model is estimated.[5]. This model

[4]Volatility clustering refers to the observation that "large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes", as defined by Mandelbrot (1963)

[5]Unsurprisingly, the GARCH(1,1) outperforms any other parameter specification in

estimates the volatility as a function of lagged volatility and squared residuals; this recursive approach is able to capture the persistence in volatility, referred to as volatility clustering. Specifically, the estimated model is the following:

$$\begin{aligned} r_{i,t} &= \mu_{i,t} + \varepsilon_{i,t} \\ \sigma^2_{i,t} &= c_2 + \alpha\sigma^2_{i,t-1} + \beta\varepsilon^2_{i,t-1} \end{aligned} \tag{1.1}$$

Different specifications of the equations above that include other controls yield similar results to the above basic formulation.

- *Volume weighted implied volatility (VW-IV):* implied volatility is defined as the value of σ that should be used into the Black-Scholes-Merton model to match the observed price of an option, obtained as the midpoint between the best closing bid price and the best closing offer price of the option. On each day, several options on the same stock are traded, whose implied volatilities are then averaged using the volume of each trade as weight.

- *Open-interest weighted implied volatility (OIW-IV):* same as before, but implied volatilities are now averaged using open interest as weight. While volume can be thought of as a proxy for liquidity, open interest may be more suited to study the impact of option expiration on the underlying stock. In fact, consider a short position in a deeply-in-the-money call option. It is likely that such an option is not traded much, as there is little uncertainty about the probability of exercising it. Hence, trading volume might be low. However, if the open interest, which measures the total number of option contracts outstanding in the market, is large, there might as well be a large trading volume in the underlying stock, as hedgers needs to continuously adjust their hedge to match the negative delta of this position. Despite the low trading volume in the option market, there certainly is an effect on the underlying stock volatility. Additionally, Chiang (2014) shows that stocks with larger open interest suffer larger negative returns on option expiration dates, so it is likely that, if any, open interest has an effect on the second moment as well.

which the orders of the GARCH are chosen by implementing the Engle LM test (see Engle (1982)) For this reason, only GARCH(1,1) estimates are reported in the paper.

Squared returns and GARCH are *empirical measures* as they are obtained from real market observations and depend on actual investors' trading behaviour. On the other hand, volume-weighted and open-interest-weighted volatilities are *implied measures* as they are inferred from the Black-Scholes-Merton model, where the option price is computed as the risk-neutral expectation of future payoff. In addition, the model assumes that the underlying stock follows a geometric Brownian motion and that volatility is constant throughout the option's life. In terms of probability measures, we can state that the former two depend on real-world probabilities, while the latter two depend on risk-neutral probabilities.

Table A1 shows the correlation among the four volatility measures described above. As they are all proxy for the second moment of the return distribution, all four measures are pairwise positively correlated. Not surprisingly, the two risk-neutral volatility proxies are strongly correlated, with a correlation coefficient higher than 80%. However, the correlation drops significantly when the empirical volatilities are considered: GARCH has a 40% correlation with the implied-volatility measures while squared returns show a correlation close to 17% with all the other measures.

To estimate the effect of option expiration on stock volatility, we run the following regression:

$$\begin{aligned}\log(\sigma^2_{i,t}) =& \alpha_i + \sum_{j=1}^{6} \beta_j \log(MON_{call,i,t})^j + \sum_{j=1}^{6} \gamma_j \log(MON_{put,i,t})^j + OI_{call,i,t} + \\ &+ OI_{put,i,t} + \delta_1 D_{Friday} + \delta_2 D_{exdate} + \varepsilon_{i,t}\end{aligned} \tag{1.2}$$

where the dependent variable is each of the four volatility measures described above and α_i represents a stock fixed effect. D_{exdate} is a dummy that is 1 on those days t which are expiration dates.[6] Hence, the coefficient δ_2 measures exactly the change in volatility on option expiration dates. However, since all expiration days are Fridays, the estimation of δ_2 might be biased by the so-called Friday effect. To account for this, the dummy D_{Friday} that takes the value of 1 if day t is a Friday is also included in the regression. Hence, δ_2 captures the impact of option expiration dates on stocks' volatility net of any

[6] Note that not all stocks have options expiring on *every* expiration date. In what follows, a separate regression accounts for stocks with no options expiring on a given expiration date.

Friday effect.
Furthermore, a sixth-order polynomial on calls and puts moneyness is included to account for the well-known volatility smile of implied volatility. Moneyness is computed as the ratio between stock price and option strike price; hence, an at-the-money call option has a moneyness equal to 1, an in-the-money call option has a moneyness larger than 1 and an out-of-the-money call option has a moneyness smaller than 1. For put options the function is exactly symmetric: the put option is in-the-money (out-of-the-money) when the underlying asset price is below (above) the strike price, hence the ratio of stock price over strike price is less (larger) than 1. This way of calculating moneyness (in comparison with the standard formula $\frac{S-K}{K}$) allows to apply the logarithmic function without dealing with negative values. Given that more than one option is traded on each stock on a specific trading day, for each option type the daily weighted averaged moneyness is computed, where the weights are the options' open interests. Allowing for calls and puts moneyness to have a different effect on volatility is necessary as exercising a call option generates a different pressure on the underlying stock market than exercising a put option. In fact, investors who are long on an in-the-money call option exercise it on an expiration day and buy a stock, but they are also likely to immediately sell it on the market to either cash in the capital gain or because they need to rebalance their portfolio. In both cases, this causes a selling pressure on the stock market which is not offset by investors who are short on in-the-money put options, as they are likely to be hedged already, nor by rebalancing of investors who are long on the put either. Evidence of this can be found in Chiang (2014), which shows that stocks with high open interest exhibit significantly negative returns on option expiration days. This effect is even stronger when the net open interest (calls minus puts) is considered.
Logarithmic function is applied to the dependent variables and the regressors so that any coefficient can be interpreted as an elasticity measure. Finally, we control for the option open interest, in line with the result of Chiang (2014). Errors are clustered at the stock level. Results of regression in Equation (1.2) are reported in Figure A1 and Table A2. Figure A1 plots the relationship between volatility and moneyness in the form of a sixth order polynomial. For each volatility measure, a comparison between call and put option moneyness is shown. In both cases, such relationship takes the typical U-shape, known as volatility smile, that we observe in financial markets since the crash of October

1987.

Numerical results are reported in Table A2. In contrast with Berument & Kiymaz (2001) which demonstrates that Friday is the most volatile day of the week, Table A2 shows that volatility is significantly lower on the last day of the week. Such difference can be easily explained by noticing that Berument & Kiymaz (2001) focuses on the S&P500 market index while this paper includes all and only the optionable stocks.

The analysis of the variable D_{exdate} is twofold: first, all the coefficients are significantly different from 0 at 5% confidence level across all measures, demonstrating that option expiration has a strong impact on volatility (besides any Friday effect). Second, the coefficients are positive for risk-neutral volatilities and negative for empirical volatilities. Such a different behaviour between realized and implied volatility around expiration could lead to profitable investment opportunities; in fact, Black-Scholes model tells us that there is a positive relationship between option price and volatility. We test the economic significance of this finding in greater details in section 1.5.

Although significant and novel, what has been described above is not necessarily the impact of option expiration on underlying stock volatility. In fact, not every stock has options expiring at every expiration date. For this reason, we split the dummy D_{exdate} into two separate dummies, namely $D_{expiring\_options}$ and $D_{no\_expiring\_options}$. The former dummy takes a value of 1 for those stocks having at least one option expiring at a given expiration date while the latter is 1 if a day is an expiration Friday but the given stock has no option expiring on that day. Table A3 shows the regression outcomes. The control variables are the same as Equation (1.2) and are not reported for easiness of reading.

Consistently with Table A2, the effect of option expiration on volatility is negative for empirical measures and positive for implied measures for stocks with at least one option expiring. We refer to this as the *direct effect.* However, we also observe an impact even on those optionable stocks which do not have any option expiring at a given expiration date. We call this the *spillover effect.* The spillover effect is positive and highly significant for risk-neutral measures only and supports the theoretical model by Danilova et al. (2018) which develops an equilibrium model in which, in the presence of option imbalance, the derivative payoff enters the household's marginal utility, hence the stochastic discount factor, affecting the pricing of all the assets in the economy.

Moreover, we argue that the two groups (stocks with expiring options and stocks with no options expiring) are equal due to: a) both groups contain optionable stocks only, so any minimum requirement necessary for being optionable is satisfied in both samples; b) any factor that varies cross-sectionally (like size, B/M, number of shares ...) is captured by the stock fixed effect which is included in the regression; c) the event “expiration date" is predetermined, exogenous and the same for every stock, so anything else happening in the market affects all the stocks in both groups in the same way. As a consequence, any difference in volatility can be only due to options reaching maturity, and we can conclude that the spillover effect is actually due to the expiration of options.

However, focusing on a single day could be misleading as rational investors are expected to predict the outcome of their option positions some days before the actual expiration. In order to have a better understanding of the dynamics of stock volatility around expiration, a regression of stock volatility on daily dummies is conducted, in which each dummy takes the value 1 for each day between two expiration dates, from 9 days before to 9 days after expiration. As above, we control for calls and puts moneyness and open interest (which are not reported for easiness of reading). A visual representation is depicted in Figure A2, which plots the two implied volatility measures against time to expiration with confidence bands at 0.5% and 99.5% levels. The plot clearly shows that volatility gradually increases as we approach the expiration day, reaching its maximum on the day before expiration. It then suddenly drops on the expiration day (day 0), when, as showed in Table A2, it is still above its average. Then volatility decreases even more in the week after expiration, and remains more or less constant for two weeks, before a further increase as the next expiration day approaches. This pattern has important consequences in terms of option prices. Such a sharp increase in implied volatility must be accompanied by an equally sharp increase in price, which will be exploited in the trading strategy discussed in section 1.5.

1.4 Possible Explanations

In this section we discuss two potential explanations for the phenomenon described above, namely delta hedging and stock pinning.

1.4.1 Delta Hedging

The first explanation hinges on the possibility that investors do not delta hedge their position when they get close to expiration. Two facts support this hypothesis: first, deeply out-of-the-money options do not bear any uncertainty so there is no reason for them to be hedged. Second, even for those investors who keep their hedge, the incentive to do so is much lower: in fact, they face a trade off between the uncertainty of experiencing a greater (if any) loss if they do not hedge, and the certainty of incurring costs (e.g. transaction costs) if they do hedge.

To test how delta hedging affects the underlying stock volatility, we construct a proxy for it using data on futures on the S&P500. In fact, investors do not delta hedge their position at the stock-level, but instead using future contracts. Using a no arbitrage argument:

$$F = Se^{r(T-t)} \quad \Rightarrow \quad \Delta \times F = \Delta \times Se^{r(T-t)} \quad \Rightarrow \quad \Delta \times Fe^{r(T-t)} = \Delta \times S$$

Hence, instead of having a position of Δ stocks, investors buy $\Delta e^{-r(T-t)}$ future contracts.

Our analysis focuses on stocks traded on the S&P500 only, whose future contracts are issued on a quarterly basis. For each trading day, we compute the delta of each stock as the weighted average of the deltas of all available options written on that stock, using the open interest as weight. Then we compute the "Delta Hedging" regressor as $\Delta F e^{-r(T-t)}$, where F is the price of the most liquid future (i.e. the one with lowest bid-ask spread) and Δ is the value-weighted delta of the stocks in the S&P500.

Table A4 shows that delta hedging have a positive and significant effect on stock volatility. In fact, higher values of delta require more trading in the underlying asset, which in turn increases the volatility. Interestingly, the first row shows that once delta hedging is accounted for, option expiration no longer has any effect on realized volatility. However, both the direct and the spillover effects on implied volatilities persist. In other words, (less) delta hedging is the reason why realized volatility drops on expiration days, while the increase in implied volatility is still unexplained.

1.4.2 Pinning

Pinning was first documented by Ni et al. (2004) which shows that around expiration dates the closing price of stocks with listed options clusters around the strike price. The authors also argue that this phenomenon is produced primarily from stocks closing around the strike on the Thursday prior to expiration and remaining in that neighbourhood on Friday, rather than from stock moving closer to the strike on the expiration Friday only. The absence of trading activity might as well justify the drop in volatility documented in the present paper.

To address this issue, on each expiration Friday and for each stock, we pick the strike price which is closest to the closing stock price and we compare it with the stock closing price from 9 days prior to 9 days after expiration. Then we count how many stocks have a price closer than \$0.5, \$0.2 and \$0.1 to the closest strike price. Figure A3 shows the distribution across expiration dates of the percentage of stocks closing near the strike around expiration; the plot supports the presence of pinning around expiration in our sample. In fact, the bottom quartile on expiration days is always above the median on any other day both before and after expiration.

The main issue is how to test for the possibility of pinning affecting volatility as we do not have a proxy for it. One possible solution is to split the sample between stocks who are likely to be subject to pinning and those who are not and, among those who are, control for the number of trades. Specifically, if the price is above the strike, the number of seller-motivated trades could be used as a proxy for pinning while, if the price is below the strike, the number of buyer-motivated trades could be used. We will address this issue in future research.

1.5 Trading Strategy

In section 1.3 we show a statistically significant increase in implied volatility around option expiration dates. Is this results also economically significant? To answer this question, we construct and test a trading strategy that exploits this finding. Figure A2 shows that implied volatility starts increasing a few days before expiration and then suddenly drops on option expiration. As in Black-Scholes model option prices increase when volatility increases, we expect

options to be over-priced throughout the expiration week. To take advantage of this, we construct a trading strategy that sells both a call and a put a few days before any expiration date. This corresponds to taking a short position on a straddle, as depicted in Figure A4. The cash flows of this strategy are: a) an inflow equal to the sum of the price of the call and the price of the put at initiation; b) a negative payoff from either the call or the put, depending on which one is exercised. The sum of the two is the net payoff of the strategy.
The outcome of this strategy strongly depends on the choice of the options to sell. First, we only keep options with bid price or open interest strictly greater than zero. These conditions ensure that investors have to pay a positive premia to buy options and that options are in positive net supply. We then pick the options depending on the following two strategies:

Strategy 1: this strategy is the most conservative as we select options based on their degree of liquidity. Intuitively, this is the strategy a very risk-averse investor, who is willing to pay a liquidity premium, will choose. Specifically:

- We first pick the options with strike closest to the current stock price, provided they are out-of-the-money. Hence, we pick the put whose strike price is closest from below to the stock price and the call whose strike price is the closest from above to the stock price;

- If there is more than one option satisfying the above condition, we select the one with lower bid-ask spread;

- If two options also have the same bid-ask spread, we choose the one with largest open interest as it is more liquid;

- Between two options with the same open interest, we select the one which is closer to maturity;

- Whenever two or more options have the same expiration date, we take the one with lowest price in the attempt of being as conservative as possible;

- Ceteris paribus, we then choose the one with higher trading volume;

- All else equal, we pick an option at random. Given the conditions above, a random choice only occurs for a couple of stock-time pairs.

Strategy 2: the second strategy simply tries to maximize the return. Differently from above, investors who choose this strategy do not care about liquidity at all; they rather try to receive the largest premia as possible. Picking the options is much simpler:

- At each point in time, we pick the out-of-the-money option with highest price. As we sell the options, we aim at maximizing the initial inflow of money;
- Among options with the same price, we pick one at random.

For each of the above strategies we compute the annualized expected excess return and the standard deviation across all the expiration dates, the Sharpe Ratio and the Information Ratio, which is computed as follows:

$$IR = \frac{\mathbb{E}[r - r_M]}{sd(r - r_M)}$$

where r_M is the return of the S&P500 index.
Panel A of Table A5 shows that both strategy 1 and strategy 2 yield positive return when implemented from 3 days to 1 day before expiration, with strategy 2 being profitable even from 4 days prior expiration. The Sharpe Ratio is increasing as we approach maturity, with an annualized value of 2.16 and 2.77 the day before for the two strategy, respectively. The Information Ratio confirms the results. The annualized Sharpe Ratio is computed by multiplying the ratio between the expected return and the standard deviation by $\sqrt{12}$ as this strategy can only be implemented 12 times a year (there are 12 expiration days in a year). The same applies to the annualized Information Ratio. In Panel B, we restrict the sample to options written on stocks in the S&P500 only. Results are extremely similar, confirming that the profitability of the strategy does not depend on small, less liquid stocks.
Table A6 tests the robustness of the results. Panel A reports the outcome of the trading strategy applied to placebo expiration dates. Specifically, each expiration date is moved by a number of trading days randomly drawn between -9 and 10. The resulting date is considered as the new expiration date. As there are at least four weeks between two consecutive expiration dates, moving them back and forward by 2 weeks ensures that placebo expiration dates do not overlap. Panel A in Table A6 shows that the conservative strategy

consistently yields negative payoffs on non-expiration dates. Regarding the aggressive strategy, although payoffs are mostly positive, the profitability is reduced significantly, with the largest Sharpe Ratio being 0.33. This could easily turn negative once transaction costs are included. Panels B and C of Table A6 mimic Panels A and B of Table A5 respectively, but restrict the sample to the period 2001-2006. The choice of this time period is justified by two reasons. First, by the expansion of the option market in the early 2000s. As a matter of fact, 76% of the options in our sample are traded from 2000 onward. As liquidity plays an important role in constructing the strategy and in managing its risk, we expect better results in the second half of our sample. In fact, the Sharpe Ratio increases from 2.15 to 2.52 for strategy 1 implemented one day before expiration. Second, from 1996 to 2001, financial markets experienced above-average volatility due to the Tech bubble. Excluding those years from the sample ensures that results are not driven by such an extreme, low-probability event.

The main drawback of this strategy is that, potentially, it can lead to extremely large negative payoff. Shorting options is risky, and shorting a straddle (or a strangle) is even riskier. For this reason, it would be interesting to assess its profitability in combination with some hedging or risk-mitigation assets. For example, a portfolio manager may want to combine the short straddle with a long low-strike put and a long high-strike call to obtain the so-called butterfly spread. Alternatively, receiver variance swaps could be used to bet on high implied volatility. In fact, the floating leg (short) is the realized underlying asset volatility, while the fixed leg (long) is usually its implied volatility.

1.6 Robustness checks

The first concern we want to address is whether the pattern in volatility is driven by the overall market sentiment rather than the option expiration as argued above. As a measure of market sentiment we use the VIX index.

For this purpose, we identify three different volatility clusters, specifically low, medium and high volatility periods. Low volatility days are those in which the VIX index is below the 5th percentile, medium volatility when VIX is between the 45th and the 55th percentile, and high volatility when the VIX is above the 95th percentile. Percentiles are computed from the time series from 1996 to 2006, consistently with the overall time frame considered so far. Table A7 shows the results of the following regression:

$$\begin{aligned}\log(\sigma_{i,t}^2) =&\alpha_i + \sum_{j=1}^{6}\beta_j \log(MON_{call,i,t})^j + \sum_{j=1}^{6}\gamma_j \log(MON_{put,i,t})^j + OI_{call,i,t}+\\ &+ OI_{put,i,t} + \delta_1 D_{friday} + \delta_2 D_{expiring\_options} + \delta_3 D_{no\_expiring\_options}+\\ &+ \delta_L D_{low\_VIX} + \delta_M D_{medium\_VIX} + \delta_H D_{high\_VIX} + \varepsilon_{i,t}\end{aligned} \tag{1.3}$$

Table A7 shows that both real-world and risk-neutral variances co-move with the VIX, as they are significantly lower when the VIX is around its lowest values ($\delta_L < 0$), significantly higher when the VIX is close to the maximum values ($\delta_H > 0$), and close to average when the VIX is also around its median value ($\delta_M \approx 0$). This is not surprising as the VIX is an average of market volatilities. Moreover, on expiration dates, stocks with at least one expiring option still exhibit lower realized volatility and higher implied volatility. Additionally, optionable stocks with no option expiring on a given expiration date also experience a rise in implied volatility. Hence, the aforementioned results are robust even when the overall market volatility is accounted for.
Even after controlling for the VIX, the relation between moneyness and volatility does not change, as it can be noticed from Figures A5 and A6 which compare the relationship between moneyness and volatility to the benchmark case as in Figure A1. No significant differences can be observed.

Second, Bollen & Whaley (2004) shows that net buying pressure positively correlates with the stock implied volatility function. The intuition is as follows: whenever there is a difference between options' demand and supply, market makers step in to absorb the imbalance and set the option price in order to receive compensation for the volatility risk or hedging costs. Under the assumption of limits to arbitrage and an upward supply curve, implied volatility will exceed actual return volatility. To control for this effect, we construct a measure for the net buying pressure as follows: a) for each stock, we define the prevailing quotes as the most frequent bid-ask pair across the daily quotes; b) trades executed at a price above (below) the prevailing bid-ask midpoint are categorized as buyer-motivated (seller-motivated) trades; c) net buying pressure is the difference between buyer-motivated and seller-motivated trades. Table A8 shows the results of the regression in which the net buying pressure

is added as a control variable.
Consistently with Bollen & Whaley (2004), net buying pressure positively correlates with implied volatility, meaning that larger differences between buyer-motivated and seller-motivated trades are associated with larger implied volatilities. However, net buying pressure does not explain the increase in volatility around expiration. In fact, coefficients on dummies of interest are still positive and significant. Both the direct and the spillover effects are robust to the inclusion of net buying pressure as a control variable.

Finally, despite the inclusion of a six-order polynomial on call and put moneyness in every regression, we acknowledge that the positive effect on implied volatility on expiration date can be caused by a reshape or a shift of the "volatility smile". To address this concern, a new volatility measure is constructed as follows: for any date-stock pair, the implied volatilities of at-the-money options are averaged, using the option open interest as weight. If no at-the-money option is available, the average is computed using the option(s) with smallest moneyness bigger than 1 and the option(s) with highest moneyness less than 1. If no options are traded on a specific date, volatility is assumed to be the same as the previous day's. In this way, we get rid of the deeply in-the-money and deeply out-of-the-money options which are typically characterized by higher volatility. This new variable is then used as dependent variable in the regression displayed in Equation (1.2). The outcome demonstrates that a significant increase in implied volatility also affects at-the-money options around expiration, so the effect of option expiration on stock volatility is not caused by a shift in the volatility smile.

1.7 Conclusion

In this paper, we first provide striking evidence that on option expiration dates realized volatility is significantly lower while implied volatility is significantly higher than their respective averages. Specifically, implied volatility builds up during the expiration week, reaching its peak the day before expiration. In addition, a similar, stronger effect is found on implied volatility of stocks with no option expiring on a given expiration date. While the decline in realized volatility can be explained by investors' delta hedging, the rise in implied volatility is a broader phenomenon which can be explained by a change in the risk-neutral measure. In particular, an increase in option imbalance around expiration could explain the finding; in fact, Danilova et al. (2018) show that, even in the presence of complete markets, option imbalance affects implied volatility, which could exhibit smile and smirk patterns.

Furthermore, we demonstrate the the impact of option expiration on stock volatility is not only statistically but also economically significant: indeed, selling a straddle a few days before expiration and holding until maturity generates a Sharpe Ratio larger than 2. Further tests demonstrate that this strategy is profitable only around expiration, but it does not depend on the time period considered.

Finally, we show that our results are robust to the inclusion of several control variables, namely the net buying pressure and the VIX index.

2. Mutual Funds' Behaviour during Financial Bubbles

2.1 Introduction

Bubble episodes are recurrent, but still partially unexplained, events in financial markets. The first known episode dates back to the XVII century, the so-called Tulip Mania, and was followed by the South Sea Bubble (1719) and the Dot-com Bubble (2000), among others. All of these episodes share some common patterns: an initial positive shock, a significant price increase (*run-up*), high trading volume and eventually a collapse. Reconciling all of them under a unique, comprehensive framework has been one of the greatest challenges in the financial literature.

In fact, in the last decades researchers have provided several explanations for the formation and evolution of bubbles in the stock market. Firstly, they tried to incorporate financial bubbles into a neoclassical model. Under the assumptions of infinite horizon and asymmetric information, some rational agents may act irrationally, causing stock prices to deviate from fundamentals (Blanchard & Watson (1983), Santos & Woodford (1997)). The drawback of these models is that the two strong assumption of infinite horizon and asymmetric information are needed to generate the price trajectory typical of financial bubbles. In fact, Tirole (1982), using a backward induction argument, argues that in a discrete-time finite-horizon setting with symmetric information stock prices cannot deviate from fundamentals. Intuitively, the presence of arbitrageurs, who act against any mispricing, would prevent any bubble from even starting. Arbitrageurs' failure to correct mispricing is at the core of a second stream of literature. The reasons for such a failure are, for example, short-sales constraints (Harrison & Kreps (1978)), heterogeneity in beliefs (Scheinkman & Xiong (2003)), and capital constraints (Shleifer & Vishny (1997)). Finally, a third explanation, pioneered by Abreu & Brunnermeier (2003), relates the evolution of bubbles to their own profitability. In fact, smart investors have an incentive to ride the bubble as long as they are able to time it, meaning they can withdraw the money before it collapses.

Building on the third explanation, this paper studies the trading activity of

mutual funds during bubble episodes and tests whether they ride and time the bubble effectively. The attention to mutual funds is justified by the crucial role that they play in financial markets.
On one hand, they are sophisticated investors who are able to quickly spot any mispricing and decide whether to correct it or exploit it. The latter theory is in line with Brunnermeier & Nagel (2004), which demonstrates that hedge funds did not exert any correcting force on stock prices during the Tech Bubble, but they were instead able to profit from the price run-up and avoid much of the downturn by withdrawing funds at the right moment. Similarly, Griffin et al. (2011) shows that most institutions were able to run the Tech Bubble until a coordinated selling effort which caused the collapse at the expenses of unsophisticated investors.
On the other hand, mutual funds differ from hedge funds in several aspects. First, their set of possible strategies is restricted by the investment objectives and risk levels stated in the prospectus, which are often linked to an index, a sector or a specific type of stocks. These constraints could potentially limit the ability of mutual funds to ride the bubble. Second, they serve different types of clients: mutual funds' money comes mainly from households, hence unsophisticated investors, while hedge funds' clients are institutions, high net-worth individuals and accredited investors, hence sophisticated investors. Such a difference is important because it is reflected in the managers' compensation: in fact, the 2-and-20 fee charged by hedge funds (which consists in 2% of management fees plus 20% of performance fees calculated on profits) creates a strong incentive to take on more risk during market booms. While the fee structure of mutual funds usually includes only a management fee as, by law, any performance fee would need to be applied equally to profits and losses. Thus, such regulation limits the risk bearing capacity of mutual funds and their incentive to ride a bubble. Third, if hedge funds take advantage of a rising market from the performance fees they charge, for mutual funds the same advantage comes from reputation. In other words, if they do not keep up with competitors' return they are subject to fund withdrawals. These in turn impact the revenues, which are calculated as a percentage of the assets under management. Indeed, one takeaway of this paper is the crucial role that households play during financial bubbles. Fourth, mutual funds represent a much larger market share, and their behaviour could affect stock prices. During the Tech Bubble, for example, mutual funds managed 7.5 trillion dollar, 60% of which was held

by equity funds only. While in 2019, they held 24% of the total outstanding shares.

This paper not only extends the analysis of investors' behaviour during financial bubbles to mutual funds, but also generalizes it in several directions. First, rather than focusing on a single bubble episode (existing literature focuses on the Tech Bubble only) it studies all the bubble episodes from the 1980s to 2018 and shows a consistent pattern across all of them. Second, rather than trades only, this paper looks at holdings and compares them with an appropriate benchmark. In fact, to prove that agents *actively* time a bubble, it is not only necessary to show that they buy stocks at the right moment, but also that they buy *more* stocks than they should or are expected to. By looking at mutual funds only, whose choices are usually restricted by the investment strategy declared in the prospectus, we can compare their trading strategies with the relative benchmark and show that they actively deviate from it. Furthermore, by looking at different holding characteristics, this paper demonstrates that mutual funds' investments predict the future return of a sector during bubble episodes. Predictability holds both in the short and medium horizon. These findings are consistent with two possible competitive stories: mutual funds *riding* the bubble versus mutual funds *creating* the bubble. In order to disentangle the two, three different tests are run and no evidence supporting an active involvement of mutual funds in generating financial bubbles is found. First, when mutual funds are aggregated on a value weighted basis, no significant over-investment is found, suggesting that the effect is not driven mainly by big funds, which are also expected to have a larger price impact. Second, there is no contemporaneous correlation between mutual funds' trades and sector abnormal returns (but there is a lead-lag positive correlation). Finally, mutual funds withdraw money 3 to 6 months before the collapse, which demonstrates that they are not directly responsible for the crash. Together with the evidence that hedge funds rode the Tech bubble (Griffin et al. (2011), showing that mutual funds are also able to ride a bubble suggests that individual investors are those who are ultimately hit by the collapse. Unfortunately, the absence of granular data on household investment holdings prevents an analysis in this direction.

Related Literature. This paper relates closely to two streams of literature: one focuses on financial bubbles and the other on mutual funds.

Regarding financial bubbles, several papers debate the topic from a theoretical perspective. Starting from rational explanations (Tirole (1982)), researches then moved to models with frictions, namely short-sales constraints (Harrison & Kreps (1978)), limits to arbitrage (Shleifer & Vishny (1997)), overconfidence (Scheinkman & Xiong (2003)), synchronization risk (Abreu & Brunnermeier (2003)) and heterogeneous beliefs (Xiong (2013)). Finally, behavioural models were introduced, in which psychological biases, such as extrapolation, play a crucial role (see, for example, Greenwood & Nagel (2009)). A complete summary of theoretical models can be found in Brunnermeier & Oehmke (2013). Such a richness of theoretical models hasn't been followed by many empirical papers. In addition, most of them focus on the Tech Bubble only. For example, Brunnermeier & Nagel (2004) studies the behaviour of hedge funds during the Tech Bubble and finds that, due to limits to arbitrage, they rode the bubble. Griffin et al. (2011) extends the analysis to other financial institutions. More recently, Liao et al. (2020) demonstrates that price and volume dynamics typical of financial bubbles can be observed in a market where investors are subject to both extrapolation and disposition effect.
The literature on mutual funds is extremely broad, but Berk & Green (2004) have a pride of place in it. The paper develops a parsimonious rational model on the relation between return and fund flows and demonstrates that persistence in returns does not reflect managers' stock picking ability. Yet, mutual fund investors chase performance, the so-called "performance based arbitrage" (see Sirri & Tufano (1998)). The inability of fund managers to consistently generate positive abnormal return after fees and transaction costs has been documented in Grinblatt & Titman (1989). Similarly, Daniel et al. (1997) show that only some mutual funds exhibit stock-picking ability, while no fund exhibits market-timing ability.

The paper is organized as follow: section 2.2 describes the data used in this project while section 2.3 reports the bubble episodes historically observed in financial markets. Mutual fund over-investment in bubble sectors is documented in section 2.4. This strategy generates a positive and statistically significant alpha. Section 2.5 demonstrates that funds' holdings have a strong predictive power and section 2.6 supports the claim that mutual funds are not responsible for the formation and the collapse of financial bubbles. Finally, section 2.7 concludes.

2.2 Data

Data on stock prices, returns and shares outstanding come from the Center of Research in Security Prices (CRSP) Monthly Stock database[1]. I restrict the sample to common stocks (code 10 or 11) and I merge CRSP data with the most recent company information available on Compustat. Each stock is associated to a specific sector using the last available SIC code from Compustat. If not available, I use the SIC code available on CRSP. Sector classification comes from Fama & French (1993) which is regularly updated on Kenneth French's website. From Kenneth's French library I also obtain the monthly risk-free rate, the market excess return and the Fama-French factors.
Data on financial institution holdings come from the Thomson Reuters Institutional Holdings database s12, which is compiled from the quarterly filings of Securities and Exchange Commission form 13F. Differently from the most common s34 database, which contains data at the fund family level, database s12 contains more granular data on *individual* mutual funds. For each fund, I have access to long positions in stocks at the quarterly level. However, no short positions and bond holdings are available. I also merge each individual fund holdings with fund characteristics from CRSP Mutual Fund database, including monthly returns, several types of fees and fund style. I use the latter to separate growth and value funds as well as to identify active funds.
Finally, from CRSP I obtain the implied volatility surface that I use to estimate the VIX and SVIX indexes. Stock-specific VIX is obtained from option data, available on OptionMetrics.

2.3 Bubble episodes

Following Greenwood et al. (2019) I identify bubble episodes as those periods in which an industry[2] experienced a value weighted return of 100% or more, both raw and in excess of market, over the past 2 years and a 100% or more raw return over the past 5 years. The choice of 100% return over two years is meant to conform to the widely accepted conviction that a bubble begins

[1]Calculated (or Derived) based on data from database name ©2018 Center for Research in Security Prices (CRSP®), The University of Chicago Booth School of Business.

[2]Industry is defined based on the 49-industry classification proposed by Fama & French (1997).

with a large price run-up. The longer 5 year horizon avoids to include in the sample those sectors which are recovering from a downturn. The first month in which these conditions are met is time 0. A bubble episode runs for 54 months, from $t = -24$ to $t = 30$. The time window is not symmetric around zero as usually the peak is reached six months after $t = 0$. In some of these episodes a crash will be eventually observed, while in other cases the cumulative return keeps growing for the whole time period considered. I define a crash as a 40% drop in price at any point in time over the 30 months after the bubble is first identified. Finally, for each bubble episode that eventually crashes, I define the dummy variable *Run-up* which takes a value of 1 from $t = -24$ to the peak. Figure B1, which shows the cumulative return of the sector *Chips*[3] in the period December 1997 - June 2002, gives a graphical representation of the timings of an episode that crashes. For sectors that do not crash, it does not make sense to construct the aforementioned dummy as the peak cannot be identified. When needed for the analysis, I will define the peak for episodes that do not crash as the average peak time of those that do crash.

Following the criteria outlined above, I identify 36 bubble episodes in the period 1980-2018, 10 of which eventually crashed. Table B1 lists the bubble episodes together with some summary statistics. Panel A reports those episodes which eventually lead to a crash in returns. As depicted in Figure B1, a bubble is defined for those sector experiencing at least 100% raw and excess return. The time of each episode (column 2 of Table B1) corresponds to the first quarter this condition is met. Column 3 contains the number of firms in each sector. Columns 4 and 5 show the cumulative return, raw and in excess of market respectively, from $t = -24$ to $t = 0$. All these numbers are above 1 by construction. Column 6 reports the percent drop in price during the crash. Finally, the last two columns show the number of months elapsing from time 0 to the peak and the crash, respectively. For example, for sector *Chips* in December 1999, the peak was reached 6 months after time 0, after the peak there was a cumulative return of 0.42, i.e. a 58% drop in price, and the crash occurred 3 months later (i.e. 9 months after time 0). Panel B reports the same information for those episodes in which the run-up was not followed by a crash. We can observe that these sectors are usually smaller in terms of number of firms, but the magnitude of the return paths, both raw and in excess of the

[3]One of the 49 sectors of the Fama-French sector classification, it includes companies producing electronic equipment. It is one of most representative sectors of the Tech bubble.

market, are very similar to those in panel A. In addition, it is worth noting that the peak occurs, on average, 7.5 months after the bubble is first identified (time 0) and the crash occurs in the next quarter. This supports the speculative explanation of financial bubbles, which would not be consistent with a long delay between the peak and the collapse. In other words, speculators won't hold on to an asset which is over-priced: they either sell it if they expect it to go back to the fundamental value or keep buying if they expect the mispricing to get worse.

2.4 Mutual fund behaviour

To study the behaviour of mutual funds during bubble episodes, it is natural to look at their holdings. For each quarter t and each sector s I compute the percentage of equity holdings that fund manager i has invested in that given sector as follows:

$$\text{Holding}_{i,s,t} = \frac{\text{Equity Holdings in} s}{\text{Total Equity Holdings}} = \frac{\sum_{j \in s} P_{j,t} X_{i,j,t}}{\sum_{j} P_{j,t} X_{i,j,t}} \tag{2.1}$$

where $X_{i,j,t}$ is the number of shares of company j that mutual fund i holds in quarter t.

However, holdings per se are meaningless if not compared with a benchmark. In fact, an increase in holdings is not necessarily an active choice of the fund manager to bet on a given sector. Alternatively, managers may change their fund composition either in response to a change in the fund's benchmark or simply due to an overall good performance of a sector which implicitly increases its market capitalization in comparison to other sector holdings. To rule out the latter explanations, I compare the holdings computed as in Equation (2.1) to the weight that a sector would have in an index made of all the stocks available on CRSP. The intuition is the following: consider a hypothetical fund manager who is agnostic about the stock market future performance. Her strategy would be to buy every stock available in the market and her portfolio would then mimic the CRSP index. Any active choice of an experienced manager would result in a deviation from the composition of the hypothetical, agnostic manager. I reckon that such a scenario, although theoretically solid, is unrealistic. Indeed, there are some stocks that managers do not invest in, like penny stocks which are usually very illiquid, hence difficult to trade.

For this reason, I refine my benchmark by considering a CRSP index which includes only stocks whose price is above \$5. Additionally, the investment set of a fund manager is limited by the fund style, which determines the ultimate objective of the fund itself. In this respect, I identify two major categories: *growth funds*, whose primary goal is to generate capital gains for investors, with no or little dividend payouts and above-average risk, and *value funds*, which instead focus on undervalued stocks and usually have a higher dividend yield. To take fund style into consideration, I identify growth funds as those which meet one of the following criteria: a) the fund name includes the word “Growth"; b) it is a growth fund according to the CRSP classification code. In my sample, 23.8% of funds can be classified as growth funds. The proper benchmark for these funds is obtained by selecting only growth stocks in the CRSP universe of stocks: following Fama & French (1993), in June of each year I sort stocks based of their book-to-market ratio and identify those in the bottom 30% as growth stocks. Following an identical procedure, I identify value funds using either the fund name or the CRSP classification code and compare their holdings with an index made of stocks in the top tercile when sorted using the book-to-market value. Approximately 12% of the funds end up in the latter category.

To test whether mutual fund managers take advantage of the run-up of the bubble, I run the following regression:

$$\text{Excess Holdings}_{i,s,t} = X_{i,t}\beta + \text{Run-up}_{s,t} + \psi_i + \sigma_s + \eta_t + \varepsilon_{i,s,t} \tag{2.2}$$

where i indicates the mutual fund, s the sector and t the quarter and ψ_i, σ_s and η_t are the corresponding fixed effects. X includes some time-varying control variables specific to each fund, namely the size of the fund, measured by the total net assets, and the return over the past quarter. Errors are clustered at the fund level. Table B2 reports the results.

In all four regressions, the coefficient of interest is positive and significant, showing that during run-ups mutual funds over-invest in bubble sectors in comparison to their benchmark. Interestingly, the behaviour of growth funds looks very similar to that of value funds, although they have a very different investment style: growth funds invest heavily in stocks which are expected to experience an increase in prices, which is exactly what happens during financial bubbles, while value funds have a more conservative approach. Apparently, value funds deviate from their objective and benefit from the bubble as well. The above regression gives a static representation on funds' strategies and confirms their ability to exploit the rapid price increase typical of bubble episodes. The following regression, by splitting the *Run-up* dummy into quarterly dummies, gives a dynamic representation of mutual funds' behaviour:

$$\text{Excess Holdings}_{i,s,t} = X_{i,t}\beta + \sum_{\tau \in B} \delta_\tau D_\tau + \eta_t + \xi_i + \sigma_s + \varepsilon_{i,s,t} \tag{2.3}$$

where B is the set of quarters of a bubble episode and D_0 is the time at which each episode reaches its peak.

Instead of reporting the coefficients δ_τ of the regression, these are depicted in Figure B2 together with their 95% confidence bands. The monotonic pattern demonstrates that mutual funds over-invest in the bubble sector during the run-up and withdraw rapidly right before the peak is reached. Afterwards, they start building up their holdings again. Figure B3 documents an interesting difference in behaviour between growth and value funds. While growth funds' capital gain oriented strategy pushes them to overweight the bubble sector from the very beginning of the run-up until two quarters before the peak, value funds are more conservative and significantly increase their exposure to bubble sectors very late, when the prices are already close to collapse.

However, a successful investment strategy cannot be determined by the manager's stock-picking ability only, but also by her market-timing ability. Especially during a bubble period, timing is the key: over-investing in a bubble sector does not necessarily involve excess returns, especially if the market has already factored the expected run-up in current prices. For this reason, in addition to a benchmark comparison, a timing analysis is needed. Options are extremely valuable in this respect. In fact, following Martin and Wagner (2019), the expected return of a stock can be obtained from option prices as follows:

$$\frac{\mathbb{E}R_{i,t+T} - R_{f,t+T}}{R_{f,t+T}} = \text{SVIX}^2_{t,T} + \frac{1}{2}(\text{SVIX}^2_{i,t,T} - \overline{\text{SVIX}}^2_{t,T})$$

where SVIX is the market SVIX, SVIX_i is the SVIX for stock i only, $\overline{\text{SVIX}}$ is the average SVIX across all available stocks and T is the time horizon considered. For all VIX-related variables, T refers to the maturity of the options used for their construction.

For each sector s, I run a cross sectional regression of the realized return on the expected return as follows:

$$\frac{R_{i,t+T} - R_{f,t+T}}{R_{f,t+T}} = \alpha + \beta \mathrm{SVIX}^2_{t,T} + \gamma(\mathrm{SVIX}^2_{i,t,T} - \overline{\mathrm{SVIX}}^2_{t,T}) + \varepsilon_i$$

The coefficient of interest is α: a positive (negative) α means that a given sector over- (under-) performed its risk-neutral expectations in the period from t to $t+T$. Given that option data are available at fixed frequencies only, the analysis could be run at 1, 3, 6 and 12 months horizons. In the interest of space, results are reported for 3 months and 6 months only.
The first test I run is whether there is a correlation between mutual fund holdings and the performance of the sector in the subsequent months. I run the following regression:

$$\hat{\alpha}_{s,t\to t+k} = X_{i,t}\beta + \text{Run-up}_{s,t} \times \text{Excess Holdings}_{i,s,t} + \eta_t + \xi_i + \sigma_s + \varepsilon_{s,t} \quad (2.4)$$

whose results are displayed in Table B3.
In the first two columns of Table B3 the whole sample is used, while in the last two columns the analysis is restricted to active funds only. For both samples, two different regressions are run: the first uses the 3-month $\hat{\alpha}$ as dependent variable, while the second uses the 6-month $\hat{\alpha}$. The first row shows that, in normal market conditions, there is a small and (almost) non significant correlation between the excess holdings and the risk-neutral excess return of a sector as measured by $\hat{\alpha}$. However, the coefficients on the interacted regressor *Run-up* × *Excess Holdings* is much stronger both in magnitude and statistical significance, showing that the stock-picking ability is enhanced during bubble episodes. Similar results are obtained when using *trades* instead of *holdings* as independent variables of the regressions. To construct fund-sector trades, I first compute stock trades as the percentage change in the number of stocks held by each mutual fund at the end of each quarter. I then aggregate them at the sector level by taking the weighted average, using the corresponding dollar holdings as weights. Regression is as follows:

$$\hat{\alpha}_{s,t\to t+k} = X_{i,t}\beta + \text{Run-up}_{s,t} \times \text{Trades}_{i,s,t} + \eta_t + \xi_i + \sigma_s + \varepsilon_{i,s,t} \quad (2.5)$$

where X also includes fund flows interacted with a stock ownership measure and a sector liquidity control (as in Lou (2012)). The former accounts for size-related constraints while the latter accounts for trading and transaction costs. Flows are computed as a percentage change in total net assets after accounting for fund mergers, while ownership is the ratio between shares held and total shares outstanding and liquidity measure is obtained following Pastor & Stambaugh (2003). In Table B4, results are even more striking. Indeed, while under normal market conditions there is a negative correlation between trades and $\hat{\alpha}$, the very same correlation flips and becomes positive during run-ups.

The results presented so far clearly demonstrate that mutual funds have the abilities (stock-picking and timing) to anticipate the run-up and benefit from it. Hence, the natural question to ask is: do they generate a positive and significant alpha from their investments? To answer this question, I first compute the value weighted return from each mutual fund's equity position at the end of each quarter. I then regress it on the Fama French factors: market excess return, size factor (SMB - Small Minus Big), value factor (HML - High Minus Low) as well as the most recent profitability (RMW - Robust Minus Weak) and investment (CMA - Conservative Minus Aggressive) factors. Results are displayed in Table B5. The first column shows the result of a standard CAPM regression, where mutual fund excess return is regressed on the market excess return and the *Run-up* dummy. The constant is omitted as it is multicollinear with the mutual fund fixed effects. The coefficient on the market excess return shows that mutual funds portfolio are well diversified and have an exposure to the market very close to 1. The coefficient of the dummy is positive and significant, showing that mutual funds earn a 1.3% abnormal return from they equity investments at a quarterly frequency. These results are robust to the inclusion of the additional factors SMB and HML as well as the more recent RMW and CMA.

2.5 Predictability

Having established that mutual funds can identify a bubble before it collapses and deviate from their usual trading strategy to take advantage of their knowledge, it is natural to check whether observing their holdings could give investors some information about the future performance of a sector. In this section, I test if mutual fund holdings can predict the future sector returns. To do so, I run the following regression:

$$R_{s,t\to t+k} = c + \beta \times Char_{s,t} + \sigma_s + \eta_t + \varepsilon_i \tag{2.6}$$

where *Char* stands for a set of mutual fund holding characteristics. The choice of these variables hinges on the findings of the previous section. In fact, we have seen that mutual funds over-weight bubble sectors compared to a hypothetical CRSP index. Hence, it is natural to use the deviation from this benchmark as the first variable that might predict future returns. Additionally, section 2.4 documents a different behaviour between growth funds and value funds, with the former being more responsive at an early stage of the bubble. Therefore, the difference in holdings between growth and value funds (henceforth GMV) could also be a powerful predicting variable. Furthermore, not only the static figures of excess holdings, but also the dynamics of holdings formation could shed some light on the future return of a sector. Dynamics can be captured by *acceleration* and *turnover*. Acceleration is defined as the change in number of shares held over the last two quarters over the change in the number of shares held in the last 4 quarters $\left(\frac{\text{Shares}_{i,t}-\text{Shares}_{i,t-2}}{\text{Shares}_{i,t}-\text{Shares}_{i,t-4}}\right)$. A value close to 1 means that the fund has significantly increased its holdings in the last 6 months or, in other words, has accelerated its buying in the recent past. Under the assumption that the fund can predict the sector performance, we expect a positive correlation between the acceleration and the future sector returns. Turnover is the percentage change in shares held $\left(\frac{\Delta\text{shares}_{i,t}}{\text{Shares}_{i,t-1}}\right)$. A higher value means that the fund is significantly increasing its holdings, which signifies that the fund expects the sector to perform well in the subsequent quarters.
For each characteristic, I run the regression in Equation (2.6) using the return over the next 3 months, 6 months and 1 year as dependent variables. As displayed in Table B8, almost all of them are able to predict future sector returns at a 5% significance level.

2.6 Endogeneity

The results presented above document the behaviour of mutual funds during bubble episodes, but do not pin down their exact role in financial markets. In fact, these findings are compatible with two very different explanations: a) mutual funds are able to predict the run-up and exploit their superior information; b) mutual funds have a price impact in the market and their simultaneous behaviour *causes* the bubble. This endogeneity problem is well known in the asset pricing literature, especially for empirical papers in which the low frequency of data (quarterly in this case) is an obstacle to any reverse causality issues. This section aims at clarifying the role played by mutual funds, showing three pieces of evidence in support of mutual funds riding the bubble rather than *directly* causing it. It is worth mentioning that this section does not exclude any involvement of mutual funds first in the run-up and then in the collapse of financial bubbles; it just demonstrates that mutual funds alone cannot generate these patterns. More likely, they are the results of the interaction between mutual fund and household investment strategies.

A value-weighted approach

First, some insights can be obtained from taking into account the size of each fund, as not all funds are big enough to have a price impact on the market. In other words, to support the claim of mutual funds causing the bubble, it must be the case that most of the over-investing found in section 2.4 comes from large funds. To test this, I aggregate individual fund data by taking the weighted average of excess holdings, using total net assets as weighting variable, and run the following regression:

$$\text{Excess Holdings}_{s,t} = X_t\beta + \text{Run-up}_{s,t} + \sigma_s + \eta_t + \varepsilon_{s,t} \tag{2.7}$$

where X now includes the sum of the total net assets of all funds at time t as well as the value-weighted return of the mutual fund industry. Results are displayed in Table B6, which shows that when the whole mutual fund industry is considered and funds are aggregated by size, the effect vanishes. This finding suggests that small mutual funds are mostly responsible for sector over-investing, therefore it is unlikely that they have a price impact.[4]

[4] Alternatively, a price impact could be generated by synchronized behaviour of several small mutual funds. Given the composition of the mutual fund industry, with the 5 largest funds commanding 46% of the market share in 2018, this is also unlikely.

Trades and abnormal returns

Second, if mutual funds had a price impact during the run-up, a positive and significant correlation between their trades and the sector abnormal return should be observed. Section 2.4 documents how funds invest in sectors which will experience positive abnormal returns in the following quarter. Using the same approach, a contemporaneous effect of trades on alpha can be analysed: if funds have a price impact, we expect a positive and significant correlation between trades and abnormal returns over the same time period. The following regression mimics Equation (2.5), with the only difference that the dependent variable α and the independent variable *Trades* are now contemporaneous. Regression output is reported in Table B7.

$$\hat{\alpha}_{s,t\to t+1} = X_{i,t}\beta + \text{Run-up}_{s,t} \times \text{Trades}_{i,s,t\to t+1} + \eta_t + \xi_i + \sigma_s + \varepsilon_{i,s,t} \tag{2.8}$$

In fact, Table B7 shows that there is no contemporaneous correlation over a 3-month horizon, suggesting that mutual fund trades have no significant impact on stock prices. The correlation is even negative over a 6-month horizon.

Comparison with other episodes

Third, some evidence can be derived from the comparison of the behaviour during two different run-ups: those of bubble episodes, which are eventually hit by a crash, versus the "fundamentally-driven" run-ups, in which the price increase is justified by the sector fundamentals and are not followed by any collapse. In this case, the focus is on the crash rather then the run-up. In fact, if mutual funds have a price impact, not only can they cause the run-up as argued above, but they might cause the collapse when they liquidate their holdings. To compare the two scenarios, the following difference-in-difference regression is run:

$$\text{Excess Holdings}_{i,s,t} = X_{i,t}\beta + \sum_{\tau\in B} \delta_\tau D_\tau D_{crash} + \eta_t + \xi_i + \sigma_s + \varepsilon_{i,s,t} \tag{2.9}$$

where D_τ is a dummy that takes into account the timing of the bubble, i.e. the distance in quarters from the peak, while D_{crash} is a dummy which is 1 for those episodes that eventually crash. Coefficients δ_τ are displayed in Figure B4. Interestingly, there is no significant difference in mutual fund behaviour during the run-up. In other words, it seems that the driver of funds' investment strategies is the future expected return, independently of the potential mispricing of the asset. In fact, during financial bubbles assets are overpriced while during "fundamentally driven" run-ups it is reasonable to say that asset are fairly priced. However, mutual funds' behaviour is not statistically different under the two scenarios. The behaviour changes when collapse approaches. For those episodes in which a crash occurs, mutual funds start reducing their exposure one quarter before the peak and two quarters before the crash. Hence, there is no evidence they are directly responsible for it. Otherwise, the crash would be contemporaneous to the withdrawal. Figure B4 also shows that after the bubble collapses, mutual funds quickly re-build their exposure: usually the market overreacts to the crash, so this is the time when new profitable investment opportunities arise.

2.7 Conclusion

In this paper, I study mutual funds' behaviour during financial bubbles and I document that mutual funds over-invest in sectors that experience a price run-up. In addition, they withdraw money before the collapse. This strategy earns a positive and significant abnormal return (4% on an annual basis) across multiple model specifications. The ability to predict future market movements is confirmed by a positive correlation between trades and future returns. More generally, a broad range of holding characteristics are proven to predict future sector returns.

These findings are consistent with the limits to arbitrage literature which explains why arbitrageurs do not exert any price correction to mispricing, but instead make profit from it. Although difficult to disentangle, exploiting and causing the bubble are two very different concepts. This paper shows that mutual funds earn abnormal profits from financial bubbles, with no evidence of them being responsible for the run-up nor the collapse.

Besides assessing the ability to time the bubble, this paper sheds some light on the fund-household relationship. Being the less sophisticated investors in the market, households are not able to predict the crash and are mostly hit by it. Although at the present date no data are available to test this hypothesis, ruling out financial institutions leaves no uncertainty on who suffers most of the losses. This is also consistent with the well-documented exacerbation of wealth inequalities as a consequence of financial bubbles. This claim has important policy implication: enhancing households' market participation could turn out to be a double-edged sword; on one hand, it is a way to achieve equality and prosperity, but on the other hand this could be detrimental to the wealth of less sophisticated investors.

3. On the Time-Series Persistence of Mutual Fund Fees

3.1 Introduction

Mutual fund fees are strongly persistent over time. A simple regression of mutual fund expense ratio on its lagged value generates an autocorrelation coefficient equal to 0.88 and, with the inclusion of a fund fixed effect, an R-squared of 99%. Not only *within* funds, but also fee dispersion *across* mutual funds is roughly constant over time. However, fund managers are sophisticated investors and should be able to calibrate their fees on investors' sentiment. Are there times when funds can charge, on average, higher expenses? More specifically, do mutual funds increase their fees during up-trending markets? Bubble episodes represent a perfect setting to answer these questions for two reasons: first, they are characterized by large price movements, during which time-varying fee policies could be investigated. Second, financial bubbles[1] are usually subject to a large media coverage, ensuring that households are well aware of the current market situation.

The present paper finds that fees increase during financial bubbles. This is true for both management fees and 12b-1 fees, which in turn generate an increase in expense ratio[2] of around 10 basis points, which corresponds to a 8.8% increase with respect to the median expense ratio. This result holds after controlling for fund characteristics as well as competition within the industry.

A possible explanation for this finding hinges on a lower investors' sensitivity to fees. In fact, in standard industrial organization settings, a lower sensitivity to fees corresponds to a steeper inverse demand function; or, in terms of quantity (i.e. market share), an increase in fees generates a smaller market share loss when fee sensitivity is lower. Additionally, media coverage might as well increase the household demand: the demand function shifts to the right, resulting in higher fees in equilibrium. However, any empirical test that estimates the sensitivity to fees suffers from endogeneity issues, due to the

[1]Specifically, the run-up component of a financial bubble

[2]Expense ratio is the annualized total cost charged to mutual fund investors.

simultaneous effect of prices on quantities and vice versa. To address this issue, an instrumental variable approach is used, where fees of a given fund are instrumented by the average fees of other funds which are not direct competitors. In the first stage I demonstrate that, despite the well-documented fee dispersion among mutual funds (see Cooper et al. (2020)), fees are correlated in the cross-section. In the second stage, I show that, while in normal times a 1% increase in fees generates a 2.9% drop in market share, a still negative but not statistically significant effect is found during financial bubbles.
However, higher fees can be justified by other factors, such as an increase in trading activity, which may result in higher transaction costs, or an increase in managers' compensation. To account for this, this paper develops a theoretical model, based on Luo (2002), which allows to split fees into two components, namely the marginal cost and the mark-up. Marginal cost is estimated as a linear function of fund characteristics, like size, age and past performance. Mark-up is strongly related to the degree of competition within the mutual fund industry. Consistently with the literature, it is measured using the Herfindal index and a normalized Herfindal index, which takes into account the number of mutual funds. These indicators are based on funds' market shares, hence they can be computed at a quarterly level. The resulting time series is then used to estimate the mark-up charged by mutual funds. Due to non-linearity of the estimation, a GMM approach is used. Additionally, as in the previous model, the estimation suffers from endogeneity. In this model, the instrumented variable is the fund size, but the intuition is similar: in fact, the size of a given fund (measured as assets under management) is instrumented using the average assets under management of competitor funds. Clearly, the target variable and the instrument are correlated, as funds operating in the same sector tend to change their size proportionally to the sector's size. Hortaçsu & Syverson (2004) provide evidence that the exclusion restriction is also satisfied. Indeed, they demonstrate that significant variation in asset under management across similar funds cannot be explained by difference in fees. A GMM is estimated both using the whole dataset and on a sample which includes bubble episodes only. While under normal market conditions the mark-up corresponds to 10% of the fees, the proportion increases to almost 20% during financial bubbles.

Related Literature. The question of whether fees *really* matter to investors has given rise to two opposite streams of literature on the mutual fund industry.

On the one hand, we have the neoclassical view, which originated from the seminal paper of Berk & Green (2004). Until then, given the apparent fierce competition between funds, combined with the inability of fund managers to outperform their passive benchmarks, it was hard to justify the high and dispersed fees charged to investors. The paper develops a model of active portfolio management that reconciles these seemingly puzzling empirical findings within a rational and competitive setting. In this model, managers increase their fees with fund size up to the point where gross alpha equals fees, and average net alpha is zero. Not surprisingly, in a competitive market investors do not earn any abnormal profit, and fees should not matter to them. The model is supported empirically by the finding in Berk & van Binsbergen (2015), which shows that, in a sample of both US and international equity funds, the average fund generates a $2m added value, but net-of-fee alphas are zero. Another key ingredient of the model is diseconomies of scale, which are also found in Pástor & Stambaugh (2012), Pástor et al. (2015), Pástor et al. (2020) and Stambaugh (2020).

On the other hand, some authors argue that mutual fund markets are not perfectly competitive, hence fees matter. For example, Hortaçsu & Syverson (2004) finds that fee dispersion among S&P500 funds can be explained by fund differentiation as well as search frictions. Similarly, Elton et al. (2004) documents that S&P500 funds, although very similar, earn very different returns. As a consequence, selecting low-fee funds outperforms the portfolio of index funds chosen by investors. This can be explained by investors' inability to arbitrage. Both papers show that mutual funds can earn abnormal profits and that after-fees alpha is not zero.

What the two streams of literature agree on, is that there is a large dispersion of fees across funds. Cooper et al. (2020) finds that, although the average net alpha is equal to zero, there is a significant and persistent cross-sectional variation in fees among mutual funds in the US, with the interquartile range of fee dispersion being around 30 basis points for S&P500 index funds. Similar dispersion have been found across countries (Khorana et al. (2008), Geranio & Zanotti (2005)). Finally, cross-sectional variation in fees correlates with fund performance: higher fees are associated with worse performance (Cooper et al. (2020), Gil-Bazo & Ruiz-Verdù (2009)), funds with front-end loads[3] have usually lower expenses (Dellva & Olson (1998)) and funds with small boards and more independent directors charge lower fees (Tufano & Sevick (1997)). Instead of looking at the cross-sectional variation, this paper focuses on the time-series variation, and shows that fees respond to market conditions. In particular, fees are significantly higher during financial bubbles. The simplest explanation is that mutual funds respond to a lower sensitivity to fees from outside investors, which is estimated using a standard empirical industrial organization model similar to Shin (2014). Additionally, the effect of such an increase in fees on mutual fund profit is estimated by calibrating a model for mutual fund competition (see Luo (2002)).

The rest of the paper is organized as follows: section 3.2 summarizes the data used, while section 3.3 provides evidence that mutual funds increase fees during financial bubbles. Robustness checks are also included. Section 3.4 develops and calibrates a model for investors' sensitivity to fees. Similarly, section 3.5 develops and calibrates, using GMM, a model to decompose fees into marginal cost and mark-up. Finally, section 3.6 concludes.

3.2 Data

Data on mutual funds come from the Center of Research in Security Prices (CRSP) Mutual Fund database[4]. It contains quarterly data on open-ended mutual funds, including a history of each mutual fund's name, investment style, asset allocation, total net assets and monthly total returns. Additionally, the fee structure and schedules of rear and front load fees are included. For each fund, I have also access to long positions in stocks at a quarterly level. Data begin at varying times between 1962 and 2008 depending on availability. From Kenneth French's Library I obtain the monthly risk-free rate, the market excess return and the Fama-French factors. I also download the updated 49-sector stock classification based on each stock SIC code.

Data on bubble episodes come from my previous work on mutual funds' investment strategy. I define a bubble episode at the sector level, provided that the following three conditions are met: a) at least 100% raw return over the

[3]Front-end loads are entry fees charged by some mutual funds.

[4]Calculated (or Derived) based on data from database name ©2019 Center for Research in Security Prices (CRSP®), The University of Chicago Booth School of Business.

past 2 years; b) at least 100% excess market return over the past 2 years; c) at least 100% return over the past 5 years. Pellicioli (2021) contains an example of bubble episode as well as the list of all price run-ups over the last 3 decades. The 49-sector classification in Fama & French (1993) is much finer than the one from the investment style information available on CRSP. The two are then matched as described in Table C1.

3.3 Fees

Fees shape the relation between funds and household: on one hand, higher fees enhance the profitability of the fund but, on the other hand, they discourage investors from providing money to the fund. From a policy perspective, fees can also be interpreted as the cost that individuals have to pay to access financial markets. Hence, understanding the fee-setting behaviour of mutual funds is interesting from both an economic and a social perspective.
There are several types of fees:

- *Management Fees*: usually calculated as a percentage of the assets under management, they are charged to compensate the fund manager for the time and effort in managing the fund. For this reason, management fees are usually higher in actively managed funds than in passive funds.

- *Front (Rear) Load:* expressed as a percentage of the amount invested, they are charged una-tantum when an individual invests in (withdraws from) the fund.

- *12b-1 Fees:* expressed as a percentage of assets under management, they are attributed to marketing and distribution costs. A maximum of 1% can be charged, with 0.75% attributable to marketing expenses and 0.25% to distribution expenses. They were originally introduced to finance marketing activities that would help the fund reach a size where economies of scales could be beneficial for the fund itself and, as a consequence, for the investors. Nowadays, they are mainly used to reward intermediaries.

- *Expense Ratio:* percentage of total investment that shareholders pay for the fund's operating expenses. As it includes waivers and reimbursements, expense ratio may be lower than management fees. Expense ratio also includes 12-b-1 fees.

Table C2 reports summary statistics of all the different types of fees. Management fees have a wide range, with a minimum value of -1.569% and a maximum value of 1.243%. In particular, management fees may turn negative due to waivers and reimbursements. The average management fees is 0.34% of the fund's net assets. Front and rear loads cannot be directly compared to management fees for two reasons: first, they are paid only once, rather than on an annual basis, hence their value should be split equally over the investors' investment horizon. Second, they usually depend on the amount invested, as they tend to be lower for low investments. The reported fees are the median values charged across different investment thresholds. More than 25% of funds do not charge entry or exit fees, although they can reach significant values of 5.75% and 2.5%, respectively. Interestingly, front load fees are larger than rear fees across the whole distribution. This can be evidence of a relatively inelastic and sticky demand for funds as investors do not mind paying entering fees while funds do not need equivalently high exit fees to retain their clients. Consistently with the legal constraint on 12b-1 fees, they range between 0 and 1%. The distribution is left skewed, with more than 25% funds charging exactly the upper-bound, i.e. 1%. Finally, in line with existing literature, the average expense ratio is slightly above 1%, with peaks of 2.5%. As it comprises all the aforementioned fees, it is a comprehensive and the most-reliable measure of the cost of the fund.

Understanding how fees change during bubble episodes is of paramount importance to analyse the fund-household relationship. In fact, there is a trade-off when mutual funds decide their fees: on one side, higher fees reduce the fund's attractiveness to outside investors, but on the other side they increase the fund's revenues. The regression in Equation (3.1), whose results are reported in Table C3, sheds some light on such a trade-off. It regresses the five types of fees on the *Run-up* dummy[5], some controls and a year fixed effect. The controls, namely the fund's net assets and return in the previous quarter, take into account the fund's performance while the fixed effect is introduced to capture the decreasing trend of fees over the last decades.

$$Fees_{i,t} = \gamma \mathit{Run\text{-}up}_t + X\beta + \eta_t + \varepsilon_{i,t} \tag{3.1}$$

[5]As defined in chapter 2, section 2.3. A graphical representation in provided in Figure B1

Table C3 shows that investing in mutual funds is significantly more costly during run-ups than in normal market conditions. In fact, even though management fees slightly decrease, the remaining fees increase significantly. As a consequence, the expense ratio is 1.3 basis points higher when the market is in a bubble.
Starting from this result, I appreciate that there are other factors affecting fund fees, namely the competition within the fund industry and the sector the fund mostly operates in. The former relates to any model of industrial organization in which, in a competitive market, revenues (fees for the mutual fund industry) should equal the marginal cost. It may be the case that, during turbulent times as financial bubbles, some funds are wiped out of the market, reducing the competition within the financial industry. This in turn allows the survivors to increase their fees. The latter is only partially related to competition. In fact, if on one hand some sectors are more heavily targeted by asset managers than others, on the other hand sectors differ for characteristics like liquidity or information transparency which affect the fees through the marginal cost. For example, transaction costs in illiquid markets are usually higher and need to be compensated by higher fees. These two determinants of fees, competition and marginal costs, will be addressed theoretically in the following sections. For now, competition can be measured empirically by looking at common holdings across funds. For example, suppose there are two funds, A and B, and three stocks, X, Y and Z. A invests 80% in X and 20% in Y, while B invests 50% in Y and 50% in Z. Stock Y is the only common stock, hence A has 20% of its holding shared with B. Note that this commonality measure is not symmetric, as B has 50% of its holdings shared with A. By considering all possible fund pairs, I can measure how "popular" A's holdings are, hence how competitive A's targeted market is. Mathematically, this measure of competition can be computed using the following equation:[6]:

$$\mathrm{MVO}_{i,t} = \frac{1}{N}\sum_{j=1}^{N}\sum_{\tau=1}^{\theta_{i,j,t}}\omega_{i,\tau,t} \qquad \text{where} \qquad \omega_{i,\tau,t} = \frac{P_{i,\tau,t}S_{i,\tau,t}}{\sum_{\gamma_i} P_{i,\gamma_i,t}S_{i,\gamma_i,t}}$$

where i and j represent the funds and τ ranges over the common holdings between i and j. Finally, S represents the number of shares, P the stock price and γ_i runs over all the stocks held by fund i. In words, $\omega_{i,\tau,t}$ is the percentage

[6]A similar measure can be found in Wahal & Wang (2011)

holdings invested by mutual fund i in stock τ at time t. For each mutual fund pair i and j, $\mathrm{MVO}_{i,t}$ sums ω across common holdings ($\tau = 1, \ldots, \theta_{i,j,t}$) and then averages across N mutual fund pairs.
Table C4 reports the outcome of regression in Equation(3.1) with the inclusion of the MVO variable as well as year-sector fixed effects. Two significant differences arise: first, the coefficient of the dummy *Run-up* is now positive for management fees, meaning that they are higher during financial bubbles. Second, entry and exit fees do not change significantly. However, the total effect on fees, measured by the expense ratio, is still positive and significant and much larger in magnitude. Expense ratios are 0.097% higher during run-ups, which represents a 10% increase when compared to its long-term, cross sectional average.

3.4 A Model of Investors' Sensitivity to Fees

This section contains a model that explains why mutual funds increase their fees during financial bubbles. In fact, there is a trade-off in raising fees: on one hand, it increases revenues for the fund, while on the other hand it discourages new inflows of money. However, an increase in fees can be rationalized by a decrease in sensitivity to fees. Does fee sensitivity actually decrease during bubble episodes? This is the questions that the subsequent model, and its calibration, answers. The model shows that, during bubble episodes, the (negative) relation between fund's share and fees decreases in magnitude: a change in fees has a significantly weaker effect on market share. In other words, the trade-off moves in favor of increasing fees, as the potential new revenues overcome the potential flows withdrawal.

3.4.1 Demand

Let $i = 1, \ldots, N$ be the set of individuals and $j = 1, \ldots, J$ be the set of mutual funds. When choosing which fund to invest in, each individual i considers some fund (observable) characteristics X_j like past return, age, size and the fees she is going to pay p_j. For simplicity, I assume that the choice of investing in fund j does not depend on the fees of other funds, p_{-j}. Other unobservable fund characteristics δ_j, like reputation and brand effects, may play a role in the individual's decision. Individuals choose which mutual fund j to invest in by maximizing their utility:

$$u_{i.j} = -\alpha p_j + X\gamma + \delta_j + \varepsilon_{i,j}$$

We can interpret the coefficient α as the sensitivity of investors to fees. Any error due to an idiosyncratic preference of individual i for mutual fund j is denoted by $\varepsilon_{i,j}$ and is assumed to follow a Type 1 Extreme Value distribution (T1EV). Under this assumption, the individual's demand for fund j is:

$$s_j(p_j) = \frac{\exp(-\alpha p_j + X\gamma + \delta_j)}{1 + \sum_{k=1}^{J} \exp(-\alpha p_k + X\gamma + \delta_k)}$$

where 1 in the denominator is a normalization of utility of a hypothetical outside option s_0.

3.4.2 Supply

Mutual funds provide the supply of investment opportunities. Given the individual's demand, they choose fees to maximise their profits given by:

$$\max_{p_j} \pi_j = \max_{p_j} s_j(p_j)A(p_j - mc_j)$$

where mc_j is the marginal cost of providing a \$1 investment and A is the total assets under management. The first order condition of the maximization problem is:

$$s_j(p_j) + \frac{\partial s_j(p_j)}{\partial p_j}(p_j - mc_j) = 0$$

The first order condition states the trade-off of an increase in fees: the first term is the benefit due to more revenues, the second term is the loss due to a lower market share s_j. In fact, the derivative of s_j is negative as an increase in fees should reduce the market shares of a fund. Specifically, we have:

$$\frac{\partial s_j(p_j)}{\partial p_j} = -\alpha s_j(1 - s_j)$$

Plugging it into the first order condition gives:

$$s_j - \alpha s_j(1 - s_j)(p_j - mc_j) = 0$$

which allows to estimate the mutual fund j's fees as:

$$p_j = mc_j + \frac{1}{\alpha(1 - s_j)}$$

Again, the intuition is quite simple: mutual funds choose their fees based on their marginal cost plus a mark-up which correlates negatively with individual's sensitivity to fees α. Hence, the change in fees can be due to either a more efficient management and/or to take advantage of a change in individuals' sensitivity. To test whether fee sensitivity changes, I estimate the model.

3.4.3 Estimation

To estimate the demand function, I regress the log odds-ratio of market share of fund j, s_j, over the outside option s_0 on fund's characteristics (quarterly return and end-of-quarter Net Asset Value), the fees it charges (expense ratio) and a fund fixed effect, as in the equation below:

$$ln(\frac{s_{j,t}}{s_{0,t}}) = \alpha \times Bubble_t \times p_j + X\gamma + \delta_j + \varepsilon_{i,j} \tag{3.2}$$

Results of the regression are displayed in Table C5. As a proxy for fees I use the expense ratio for the most recent completed fiscal year, measured as ratio of total investment that shareholders pay for the fund's operating expenses, including 12b-1 fees, waivers and reimbursements. In fact, expense ratio is what investors actually pay as a result of a bargaining power with the fund and it is the variable they take into consideration when they choose among several funds. $Bubble_t$ is a dummy that takes a value of 1 if at time t the market experiences a bubble episode.

As in the model above, under normal market conditions the coefficient of fees is negative and significant, meaning that, coeteris paribus, funds with higher fees have a lower market share. However, the interacted coefficient is positive: during bubble episodes, the sensitivity of fees, which is the sum of both coefficients, becomes less negative. In other words, the sensitivity of fees decreases significantly (in absolute value). During normal times, the regression coefficient of market share on fees is -0.551, as reported in Table C5. However, during bubble episodes, the same coefficient decreases in magnitude by 0.359 (corresponding to 65%) to -0.192. This simple model explains clearly the trade-off of an increase in mutual fund fees but suffers from two limitations: a) the coefficients of the regression of market share (quantity) on fees (price) is biased due to simultaneity/reverse causality; b) it does not disentangle the marginal cost and the mark-up, making it unclear whether higher fees are needed because trading is more costly or they are simply a transfer from household to fund managers.[7] The first concern is addressed with a different regression approach, while the second is discussed in the next section.

From an econometric perspective, reverse causality issues can be resolved with an instrumental variable regression. In particular, the instrument needs to possess two characteristics: a) it must correlate with the endogenous regressor; b) it must affect the dependent variable through the endogenous regressor only (the so-called exclusion restriction). I argue that instrumenting the expense ratio of fund i with the average expense ratio of all other funds j belonging to a different classification group than fund i satisfies both the above-mentioned properties. The first condition is ensured by the long-term downward trend of fees within the mutual fund industry over the last three decades and by their contemporaneous correlation with the market performance. I argue that the second condition holds as investors first choose the sector they want to invest in and then they pick the cheapest fund within that sector. Hence, the market share of fund i should not be directly correlated with the fees of funds operating in other sectors. Mathematically, we can decompose the percent market share of fund i operating in sector S as follows:

$$\frac{TNA_i}{\sum_j TNA_j} = \frac{TNA_i}{\sum_{s \in STNA_s}} \frac{\sum_{s \in STNA_s}}{\sum_j TNA_j}$$

The first term compares the size of fund i with the size of sector S, which is independent of fees charged by funds investing in other sectors. The second term is the ratio of the size of sector S with the size of the mutual fund industry, which again is independent of fees as the investors' first choice is which sector

[7]Using the fitted value of the regression in Equation (3.2), $\hat{s}$, and the estimated $\hat{\alpha}$, I can estimate the marginal cost and the mark-up as:

$$mc_j = p_j - \frac{1}{\hat{\alpha}(1-\hat{s}_j)} \qquad \text{and} \qquad mark\text{-}up_j = \frac{1}{\hat{\alpha}(1-\hat{s}_j)}$$

However, market share is very persistent and it is unlikely that it immediately incorporates the effect of a change in fees. For this reason, the estimation of the mark-up is discussed in section 3.5.

to invest in (independently of fees). This is particularly true during financial bubbles. Suppose, for example, that the Technology sector is experiencing a financial bubble. Even if fees of some other sectors, let's say Utilities, are significantly lower, investors will invest in Tech funds to gain from the price run-up. Table C6 shows the results of the IV regressions. In the first stage, expense ratio is regressed on the average expense ratio of non-competitor funds, controlling for size, performance and adding sector and year fixed effects. Due to the strong persistence of fees within funds, a fund fixed effect is not added. In fact, the inclusion of a fund fixed effect would generate an R-squared higher than 96%. The first stage fitted values would then be extremely close to the variable they are instrumenting, making the whole IV regression meaningless. In the second stage, the log market share is regressed on the first stage fitted values with the same controls.

The first two columns of Table C6 reports the first and second stage regressions, respectively, for the time-sector pairs that do not constitute a financial bubble. The first regression shows that there is a strong correlation between the expense ratio of fund i and those of non-competitor funds, which proves that the instrument is sufficiently strong. The second regression shows that, in normal times, there is a negative sensitivity between fees and market share. A 1bp increase in fees generates a -2.9% decrease in market share. In the remaining two columns, the sample is restricted to include those time-sector pairs which represents price run-ups. The first stage still indicates that the chosen instrument in strong. However, the second stage shows that investors' sensitivity to fees is still negative (as any inverse demand function), but no longer significant. In other words, investors do not necessarily withdraw money from expensive funds, provided that they generate sufficiently large returns. An alternative specification of the model uses the change in market share as dependent variable in the second stage. The intuition is straightforward: rather than the market share level, fees are more likely to influence the fund flows, hence the change in market share. An increase (decrease) in fees should generate an outflow (inflow) of money and a negative (positive) change in market share. Table C7 shows the IV two stage regression using the logarithm of change in market share as a dependent variable. Results are both qualitatively and quantitatively very similar.

3.5 Marginal Cost vs. Mark-Up

Instead of looking at the fund-investor relation, this section estimates the mark-up component of the fees charged by mutual funds. Intuitively, funds are able to charge higher mark-ups when investors' sensitivity to fees is lower. This paper does not investigate the determinants of a lower elasticity to fees, although funds' superior information and easier access to financial markets are reasonable explanations. Based on data availability, the following model is a simplified version of Luo (2002).

3.5.1 Marginal Cost

Following Tufano & Sevick (1997) and Malhotra & McLeod (1997), mutual fund costs are assumed to be function of both fund characteristics and performance related variables. Among the former, total net assets ($A_{i,t}$) and fund age are used. It is widely recognized in the literature the existence of economies of scale. Hence, larger funds are expected to have lower marginal costs. Mathematically, there are economies of scale when the cost function is not linear in the quantity, thus its first order derivative (i.e. the marginal cost) is function of the quantity as well. The second fund feature considered, $Age_{i,t}$, accounts for unobservable fund manager characteristics, like experience and reputation. Fund managers who have been investing in financial markets for several years have surely more experience and, assuming they have been profitable, they also have an above-average reputation. Assuming there is a learning curve effect, the older the fund is, the cheaper it is to manage. Regarding performance related variables, quarterly return is included among the marginal cost determinants. In fact, a well-performing fund is usually more costly to manage than a fund with lower profitability, due to pre-trading analysis, risk management and managers' compensation, among others. To avoid any contemporaneous effect between return and fees, which could bias the estimation, I use the lagged return. In fact, on one hand, the return time series displays positive autocorrelation, making lagged return a good independent variable. On the other hand, this ensures there is no reverse causality as fees in quarter t could not have affected the fund's performance in the previous quarter.

In summary, fund marginal cost is modelled as follows:

$$MC_{i,t} = f(A_{i,t}, Age_{i,t}, Ret^{Q}_{i,t-1}) \tag{3.3}$$

where $A_{i,t}$ denotes the assets under management and $Ret^{Q}_{i,t-1}$ the quarterly return of fund i at time $t-1$.

3.5.2 Demand function

For the purpose of this paper, there is no need to explicitly model the demand of funds ($q_{i,t}$) from households. However, some considerations are still necessary. The first choice that investors make is which sector to invest their money into. Hence, it is reasonable to model the inverse demand function as function of the sector size, denoted by $A^{S}_{i,t}$. In addition, demand depends heavily on past performance, usually known as *performance based arbitrage*[8]. According to this hypothesis, outside investors do not understand the trading strategies employed by fund managers and provide additional money following good performance and withdraw funds following past losses. Finally, demand depends on fund's reputation and experience, which are unobservable. A good proxy for them is age, under the assumption that long lasting funds have a good record of past return, hence a good reputation, and clearly more experience. In summary, when deciding which sector to invest in, outside investors take into account the size of the sector as well as the performance and the age of the fund. The (inverse) demand function of fund i at time t can be written as:

$$p_{i,t} = \Phi^{S}(A^{S}_{i,t}, \mu_{i,t}, Age_{i,t})$$

where $A^{S}_{i,t}$ is total net assets of mutual funds in sector S and $\mu_{i,t}$ is the average annual return over the last three years.

3.5.3 Utility maximization

In the previous model, mutual funds choose their fees to maximise their profit. In this model, they still maximise their profit but, using the inverse demand

[8] See Shleifer & Vishny (1997)

function, the choice variable is now A rather p. That is:

$$\max_{A_{i,t}} \quad [\Phi^S(A^S_{i,t}, \mu_{i,t}, Age_{i,t})A_{i,t} - C(\cdot)]$$

Taking first order conditions:

$$\frac{\partial \Phi^S}{\partial A^S_i}\frac{\partial A^S_i}{\partial A_i}A_i + \Phi^S(\cdot) - \frac{\partial C(\cdot)}{\partial A_i} = 0$$

where time subscripts have been removed to make the notation easier. Rearranging term and using $p_i = \Phi^S(\cdot)$:

$$p_i\left(1 + \left(\frac{\partial \Phi^S}{\partial A^S_i}\frac{A^S}{p_i}\right)\left(\frac{\partial A^S_i}{\partial A_i}\frac{A_i}{A_s}\right)\right) - MC(\cdot) = 0 \tag{3.4}$$

Note that $(\partial \Phi^S/\partial A^S_i)(A^S/p_i)$ is the negative of the inverse elasticity of demand. Let's define:

$$-\frac{1}{\eta^S_i} = \frac{\partial \Phi^S}{\partial A^S_i}\frac{A^S}{p_i}$$

Similarly, $(\partial A^S_i/\partial A_i)(A_i/A_s)$ is the elasticity of sector S total fund with respect to the total assets of fund i. Let's define:

$$\delta^S_i = \frac{\partial A^S_i}{\partial A_i}\frac{A_i}{A_s}$$

Substituting the last two equations into Equation (3.4), the first order condition can be rewritten as:

$$p_i\left(1 - \frac{\delta^S_i}{\eta^S_i}\right) = MC(\cdot) \qquad \text{or} \qquad p_i = MC(\cdot)\left(1 - \frac{\delta^S_i}{\eta^S_i}\right)^{-1} \tag{3.5}$$

The above equation decomposes the optimal fees into two components, namely the marginal cost and the mark-up. In case of perfect competition, an infinitesimal increase in fees of fund i (p_i) causes investors to withdraw their money and invest into cheaper funds. In other words, the elasticity of demand is infinite ($\eta^S_i \to \infty$). Equation (3.5) becomes $p_i = MC(\cdot)$. Fees equal marginal cost, and there is no room for mark-up. Similarly, in case of perfect competition within a sector, investors would easily transfer money across competitors without changing the total asset of the sector as a whole. Hence, $\partial A^S_i/\partial A_i = 0$ and $\delta^S_i = 0$. Again, fees equal marginal cost and investors do not pay for any

mark-up. In all the other cases, both the demand elasticity (η_i^S) and the sector elasticity (δ_i^S) are positive and fees exceed marginal cost. Furthermore, it is likely that the competition across mutual funds within each sector changes over time. For example, during the run-up of a bubble, more funds may want to reap the benefit of a rapidly growing market, increasing the competition. However, when the price is close to its peak and riding the bubble has become increasingly risky, some funds may exit the market, reducing the competition. A few months after the collapse (usually between 3 to 6 months), when prices have dropped below their fundamental value, competition may restore as new profitable investment opportunities arise.
To test for the existence of a mark-up, an explicit expression of the term $(1 - \delta_i^S/\eta_i^S)^{-1}$ is needed. Economic literature agrees on using the Herfindal index as a measure of competition. It is defined as the sum of squared market shares within an industry. When market shares are calculated as fractions, it ranges between $1/N$ and 1, where $1/N$ represents a perfectly competitive market while 1 indicates a monopoly. For the mutual fund industry, the market share is measured using the total net assets (A); hence the Herfindal index for sector S at time t is defined as:

$$h_t(S) = \sum_{i \in S} s_i^2 = \sum_{i \in S} \left(\frac{A_i}{\sum_{j \in S} A_j} \right)^2 \tag{3.6}$$

One limitation of this indicator is that it depends on the number of firms within each sector. Consider for example two sectors, A and B, with equally distributed market share. In A operates N_A funds, while in B operates N_B funds. The market shares in the two sectors are $1/N_A$ and $1/N_B$ respectively and, using Equation (3.6), $h(A) = 1/N_A$ and $h(B) = 1/N_B$. Even if in both sectors market shares are equally distributed, the Herfindal indexes are different. In other words, the Herfindal index is affected by the number of funds. However, Sutton (1991) shows that, in many oligopolistic markets, there is no relation between the number of participants and the mark-up. Hence, a normalization of the Herfindal index may be needed for an accurate estimation. The normalization is as follows:

$$h_t^N(S) = \frac{h_t(S)}{h_t^C(S)} - 1$$

where $h_t^C(S) = 1/N_S$ is the Herfindal index under equally distributed market shares. Going back to the example above, $h(A) = 1/N_A$ and $h^C(A) = 1/N_A$, hence $h^N(A) = 0$. Similarly, $h(B) = 1/N_B$ and $h^C(B) = 1/N_B$, hence $h^N(B) = 0$. The normalized Herfindal index for sector S is bounded between 0 and $N_S - 1$ and decreases with competition. Although they have a different interpretations, both the Herfindal index and the normalized Herfindal index are able to capture the degree of competitiveness which explains the mark-up embedded in the mutual fund fees. Hence, I model the mark-up as function of both, as in the equation below:

$$\left(1 - \frac{\delta_i^S}{\eta_i^S}\right)^{-1} = 1 + \theta_1 h_t(S) + \theta_2 h_t^N(S) \tag{3.7}$$

As discussed above, note that when $h(S) \to 0$ and $h^N(S) = 0$ then $(1 - \delta_i^S/\eta_i^S)^{-1} \to 1$, hence there is no mark-up. As the market becomes less and less competitive, $h(S)$ and $h^N(S)$ increase and the mark-up becomes positive. Substituting Equation (3.7) into Equation (3.5), we obtain:

$$p_i = MC_i(\cdot)(1 + \theta_1 h_t(S) + \theta_2 h_t^N(S)) \tag{3.8}$$

3.5.4 Estimation

To estimate the model, we need an explicit formula for the marginal cost described in Equation (3.3). For simplicity, let's assume that the marginal cost is linear in the variable described above; hence:

$$MC_{i,t} = \gamma_1 A_{i,t} + \gamma_2 Age_{i,t} + \gamma_3 Ret_{i,t-1}^Q$$

Substituting into Equation (3.8), it follows that fees can be expressed as:

$$p_i = (\gamma_1 A_{i,t} + \gamma_2 Age_{i,t} + \gamma_3 Ret_{i,t}^Q)(1 + \theta_1 h_t(S) + \theta_2 h_t^N(S)) \tag{3.9}$$

Compared to the model in section 3.4, this formulation allows to estimate the mark-up as function on the Herfindal indexes and the parameters $\hat{h}(S)$ and $\hat{h}^N(S)$. However, it is still subject to the endogeneity problem caused by the presence of asset under management among the explanatory variables. Similarly to the previous section, this issue can be solved using an instrumental variable. I claim that the average assets under management of competitor funds ($A_{-i,t}$) is a good instrument. For fund i operating in sector S, I define $A_{-i,t}$ as follows:

$$A_{-i,t} = \frac{1}{N_s - 1} \sum_{j \neq i} A_{j,t} \tag{3.10}$$

The variable $A_{-i,t}$ satisfies both the requirements of a good instrumental variable. In fact, it correlates with $A_{i,t}$ as both fund i and the competitors operate in the same sector at the same time and must have similar holdings. An increase in the market value of a sector positively affects all the funds which invest in it, including fund i. To address the exclusion restriction, I refer to Hortaçsu & Syverson (2004) which finds that significant variation in asset under management across similar funds cannot be explained by difference in fees. Hence, assets under management of competitors have an effect on fees of fund i only through i's assets under management.

As the model is non-linear, it is estimated using the generalized method of moments (GMM). Let X be the set of explanatory variables in Equation (3.9) and define ε as the unexplained component of fees as follows:

$$\varepsilon_{i,t} = p_{i,t} - MC(X_{i,t})(1 + \theta_1 h_t(S) + \theta_2 h_t^N(S))$$

Let Z be the set of instruments, which includes $Age_{i,t}$, $Ret^Q_{i,t-1}$, $h_t(S)$ and $h_t^N(S)$. In addition, I include $A_{-i,t}$ as instrument for total net assets as well as $A^2_{-i,t}$[9]. The GMM minimizes the following quadratic function:

$$G(\gamma, \theta; Z)' W G(\gamma, \theta; Z)$$

where γ and θ are the parameter vectors and W is the weighting matrix (no restrictions are imposed on W). The matrix G contains the sample moment conditions $\mathbb{E}[Z'\varepsilon] = 0$, as, by the exclusion restriction, all variables in Z are now independent to ε.

The model is estimated both using all available information and after restricting the sample to bubble episodes only. Results are reported in Table C8.

Consistently with the literature, there is a negative correlation ($\gamma_1 < 0$) between fees and funds' total net assets. In fact, the presence of economies of scales reduces the marginal cost as the fund's size increases. The coefficient γ_2 shows that "older" funds charge higher fees; given that age is a proxy for

[9] In all the estimates that follow, the inclusion of an additional instrument does not violate the J-test for over-identifying restrictions.

reputation and expertise, it is reasonable that experienced funds are rewarded more. The coefficient γ_3 supports the performance based arbitrage assumption: funds with higher past return attracts more investors, hence they tend to be more expensive. Finally, θ_1 shows that fees increase with the Herfindal index (hence decrease with competition), with a minor correction in highly populated sectors as represented by the normalized Herfindal index (θ_2). Similar results are obtained during bubble episodes, except for the negative coefficient γ_3, which requires further analysis. Surely, during financial bubbles all funds experience particularly large returns, which makes the return-fee relation difficult to interpret.

The main advantage of this model is that it allows to estimate the mark-up that mutual funds charge, expressed as a percentage of the fees. Let's define the mark-up as:

$$\text{Mark-up}(\%) = \frac{\text{Fees} - \text{Marginal Cost}}{\text{Fees}}$$

Substituting Equation (3.8) yields:

$$\text{Mark-up}(\%) = 1 - \frac{1}{1 + \theta_1 h_t(S) + \theta_2 h_t^N(S)}$$

Substituting the estimate $\hat{\theta}_1$ and $\hat{\theta}_2$ and averaging over time, I can compute the average mark-up. Across the whole sample, 88.39% of fees is used to cover marginal costs, while the remaining 11.61% represents the fund's mark-up. These percentages change dramatically during financial bubbles, when the mark-up almost doubles to 21.5% of fees.

3.6 Conclusion

Despite the persistence of fees over time, this paper documents a statistically significant increase of mutual fund fees during financial bubbles. A model on the relation between mutual funds and outside investors demonstrates that such an increase can be explained by a lower sensitivity to fees of households. Furthermore, a second theoretical model shows that higher fees are also associated to a larger percent mark-up that mutual funds charge on top of their marginal cost.

In light of the recent development of several behavioural models that reconcile the formation of financial bubbles with the extrapolative behaviour of households, a deeper analysis of fees could shed some light on the relation between mutual funds and outside investors. Specifically, Barberis et al. (2018) present a model where the interaction between fundamental traders and extrapolators could generate a price pattern that resembles those of financial bubbles. Showing that the increase in fees is larger than what the change in sensitivity would explain, could demonstrate some rent extraction at the expenses of unsophisticated investors. This would also result in an enhanced profitability of funds during financial bubbles. Although the results in this paper point in this direction, there is no evidence of any wealth transfer from naive households to smart fund managers. Similarly, the increase in mark-up does not necessarily mean that mutual funds make abnormal profit, after risk is factored in. A deeper analysis of these of these topics is left for future research.

Bibliography

Abreu, D. & Brunnermeier, M. K. (2003). Bubbles and crashes. *Econometrica*, 71, 173–204.

Avellaneda, M., Kasyan, G., & Lipkin, M. D. (2012). Mathematical models for stock pinning near option expiration dates. *Communications on Pure and Applied Mathematics*, 65(7), 949–974.

Bansal, V. K., Pruitt, S., & Wei, K. C. J. (1989). An empirical reexamination of the impact of cboe option initiation on the volatility and trading volume of the underlying equaties 19731986. *The Financial Review*, 24(1), 19–29.

Barberis, N., Greenwood, R., Jin, L., & Shleifer, A. (2018). Extrapolation and bubbles. *Journal of Financial Economics*, 129(2), 203–227.

Bartunek, K. S. (1996). Option delisting of stocks that continue trading: An examination of welfare effects. *Financial Review*, 31(3), 565 – 583.

Berk, J. & Green, R. (2004). Mutual fund flows and performance in rational markets. *Journal of Political Economy*, 112, 1269–1295.

Berk, J. B. & van Binsbergen, J. H. (2015). Measuring skill in the mutual fund industry. *Journal of Financial Economics*, 118(1), 1–20.

Berument, H. & Kiymaz, H. (2001). The day of the week effect on stock market volatility. *Journal of Economics and Finance*, 25(2), 181–193.

Blanchard, O. J. & Watson, M. W. (1983). Bubbles, rational expectations and financial markets. *National Bureau of Economic Research*.

Bollen, N. P. B. & Whaley, R. E. (2004). Does net buying pressure affect the shape of implied volatility functions? *Journal of Finance*, 59(2), 711–753.

Brunnermeier, M. K. & Nagel, S. (2004). Hedge funds and the technology bubble. *The Journal of Finance*, 59, 2013–2040.

Brunnermeier, M. K. & Oehmke, M. (2013). *Bubbles, Financial Crises, and Systemic Risk*, (pp. 1221–1288). Elsevier.

Chiang, C.-H. (2014). Stock returns on option expiration dates: Price impact of liquidity trading. *Journal of Empirical Finance*, 28(Supplement C), 273 – 290.

Chiang, C.-H. (2017). Trading volume and option expiration date. *Working Paper*.

Christensen, B. J. & Prabhala, N. R. (1998). The relation between implied and realized volatility. *Journal of Financial Economics*, 50(2), 125–150.

Conrad, J. (1989). The price effect of option introduction. *The Journal of Finance*, 44(2), 487–498.

Cooper, M. J., Halling, M., & Yang, W. (2020). The persistence of fee dispersion among mutual funds. *Review of Finance*.

Daniel, K., Grinblatt, M., Titman, S., & Wermers, R. (1997). Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of Finance*, 52(3), 1035–1058.

Danilova, A., Julliard, C., & Stoev, Y. (2018). Speculative trading and derivative market imbalances. *Working Paper*.

Dellva, W. L. & Olson, G. T. (1998). The relationship between mutual fund fees and expenses and their effects on performance. *Financial Review*, 33(1), 85–104.

Detemple, J. & Jorion, P. (1990). Option listing and stock returns: An empirical analysis. *Journal of Banking and Finance*, 14(4), 781–801.

Elton, E. J., Gruber, M. J., & Busse, J. A. (2004). Are investors rational? choices among index funds. *The Journal of Finance*, 59(1), 261–288.

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4), 987–1007.

Fama, E. F. & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3 – 56.

Fama, E. F. & French, K. R. (1997). Industry costs of equity. *Journal of Financial Economics*, 43(2), 153 – 193.

Fama, E. F. & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1 – 22.

French, K. R. & Roll, R. (1986). Stock return variances: The arrival of information and the reaction of traders. *Journal of Financial Economics*, 17(1), 5 – 26.

Geranio, M. & Zanotti, G. (2005). Can mutual funds characteristics explain fees? *Journal of Multinational Financial Management*, 15(4), 354–376.

Gil-Bazo, J. & Ruiz-Verdù, P. (2009). The relation between price and performance in the mutual fund industry. *The Journal of Finance*, 64(5), 2153–2183.

Golez, B. & Jackwerth, J. (2012). Pinning in the s&p 500 futures. *Journal of Financial Economics*, 106(3), 566–585.

Greenwood, R. & Nagel, S. (2009). Inexperienced investors and bubbles. *Journal of Financial Economics*, 93(2), 239–258.

Greenwood, R., Shleifer, A., & You, Y. (2019). Bubbles for fama. *Journal of Financial Economics*, 131(1), 20–43.

Griffin, J. M., Harris, J. H., Shu, T., & Topaloglu, S. (2011). Who drove and burst the tech bubble? *The Journal of Finance*, 66(4), 1251–1290.

Grinblatt, M. & Titman, S. (1989). Mutual fund performance: An analysis of quarterly portfolio holdings. *The Journal of Business*, 62(3), 393–416.

Harrison, J. M. & Kreps, D. (1978). Speculative investor behavior in a stock market with heterogeneous expectations. *The Quarterly Journal of Economics*, 92(2), 323–336.

Hortaçsu, A. & Syverson, C. (2004). Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds*. *The Quarterly Journal of Economics*, 119(2), 403–456.

Hu, J. (2014). Does option trading convey stock price information? *Journal of Financial Economics*, 111(3), 625 – 645.

Khorana, A., Servaes, H., & Tufano, P. (2008). Mutual Fund Fees Around the World. *The Review of Financial Studies*, 22(3), 1279–1310.

Klemkosky, R. C. (1978). The impact of option expirations on stock prices. *Journal of Financial and Quantitative Analysis*, 13(3), 507 – 518.

Klemkowsky, R. C. & Maness, T. S. (1980). The impact of options on the underlying securities. *The Journal of Portfolio Management*, 6(2), 12 – 18.

Koopman, S. J., Jungbacker, B., & Hol, E. (2005). Forecasting daily variability of the s&p 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance*, 12(3), 445 – 475.

Lakonishok, J., Lee, I., D. Pearson, N., & M. Poteshman, A. (2006). Option market activity. *Review of Financial Studies*, 20, 813–857.

Liao, J., Peng, C., & Zhu, N. (2020). Price and volume dynamics in bubbles. *Working Paper*.

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

Luo, G. Y. (2002). Mutual fund fee setting, market structure and mark-ups. *Economica*, 69, 245–271.

Malhotra, D. K. & McLeod, R. W. (1997). An empirical analysis of mutual fund expenses. *Journal of Financial Research*, 20(2), 175–190.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *The Journal of Business*, 36, 394–394.

Mayhew, S. & Mihov, V. (2005). Short sale constraints, overvaluation, and the introduction of options. *Working Paper, presented at 205 AFA Philadelphia Meeting*.

Ni, S. X., Pearson, N. D., & Poteshman, A. M. (2004). Stock price clustering on option expiration dates. *Journal of Financial Economics*, 78(1), 49 – 87.

Pastor, L. & Stambaugh, R. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, 111(3), 642–685.

Pellicioli, A. (2021). Mutual funds' behaviour during financial bubbles. *Working paper*.

Pástor, L. & Stambaugh, R. F. (2012). On the size of the active management industry. *Journal of Political Economy*, 120(4), 740–781.

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

Pástor, L., Stambaugh, R. F., & Taylor, L. A. (2020). Fund tradeoffs. *Journal of Financial Economics*, 138(3), 614–634.

Santos, M. S. & Woodford, M. (1997). Rational asset pricing bubbles. *Econometrica*, 65(1), 19–57.

Scheinkman, J. & Xiong, W. (2003). Overconfidence and speculative bubbles. *Journal of Political Economy*, 111(6), 1183–1219.

Schwert, G. W. (1989). Why does stock market volatility change over time? *The Journal of Finance*, 44(5), 1115 – 1153.

Shin, C. H. (2014). *An Industrial Organization Approach to International Portfolio Diversification: Evidence from the U.S. Mutual Fund Families*. Finance and Economics Discussion Series 2014-78, Board of Governors of the Federal Reserve System (U.S.).

Shleifer, A. & Vishny, R. W. (1997). The limits of arbitrage. *The Journal of Finance*, 52(1), 35–55.

Sirri, E. R. & Tufano, P. (1998). Costly search and mutual fund flows. *The Journal of Finance*, 53(5), 1589–1622.

Skinner, D. J. (1989). Options markets and stock return volatility. *Journal of Financial Economics*, 23(1), 61–78.

Sorescu, S. M. (2000). The effect of options on stock prices: 1973 to 1995. *The Journal of Finance*, 55(1), 487–514.

Stambaugh, R. F. (2020). Skill and profit in active management. *Working paper*.

Sutton, J. (1991). *Sunk costs and market structure*. MIT Press.

Tirole, J. (1982). On the possibility of speculation under rational expectations. *Econometrica*, 50(5), 1163–81.

Trennepohl, G. L. & Dukes, W. P. (1979). Return and risk from listed option investments. *Journal of Financial Research*, 2(1), 37 – 49.

Tufano, P. & Sevick, M. (1997). Board structure and fee-setting in the u.s. mutual fund industry. *Journal of Financial Economics*, 46(3), 321–355.

Wahal, S. & Wang, A. Y. (2011). Competition among mutual funds. *Journal of Financial Economics*, 99(1), 40 – 59.

Whiteside, M. M., Dukes, W. P., & Dunne, P. M. (1983). Short term impact of option trading on underlying securities. *Journal of Financial Research*, 6(4), 313 – 321.

Xiong, W. (2013). *Bubbles, Crises, and Heterogeneous Beliefs*, (pp. 663–713). Cambridge University Press.

A. Appendix to Stock Volatility on Option Expiration Dates

Table A1: Correlation among Volatility Measures

This Table reports the pairwise correlation between the four volatility measures. For each pair, correlations are calculated using only those days in which both measures are available. As all four measures are proxies for the second central moment of the return distribution, they are all positively correlated with each other.

	Squared return	GARCH(1,1)	OIW var	VW var
Squared return	1	0.160	0.169	0.176
GARCH(1,1)	0.160	1	0.385	0.415
OIW var	0.169	0.385	1	0.804
VW var	0.176	0.415	0.804	1

Table A2: Regression of Volatility on Option Expiration Dates

This Table reports the outcome of the regression in Equation (1.2). Only the two coefficients of interest are reported. Each regression includes a stock fixed effect and errors are clustered at the stock level.

	Dependent variable:			
	$\text{Log}(\sigma^2)$			
	Squared return	GARCH(1,1)	VW-IV	OIW-IV
	(1)	(2)	(3)	(4)
D(Friday)	-0.007^{***}	0.009^{***}	-0.042^{***}	-0.032^{***}
	(0.0004)	(0.001)	(0.001)	(0.001)
D(expiration date)	-0.002^{***}	-0.008^{***}	0.032^{***}	0.035^{***}
	(0.001)	(0.002)	(0.002)	(0.001)
Stock FE	Y	Y	Y	Y
Observations	2,128,988	1,545,786	2,128,988	2,128,988
R^2	0.012	0.059	0.079	0.089
Adjusted R^2	0.012	0.058	0.079	0.089
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure A1: Volatility as a Function of Moneyness

This figure plots the six-order polynomial of the regression of stock volatility on option moneyness, as in Equation (1.2). Moneyness is computed as $(S-K)/K$. For each volatility measure, call and put moneyness are compared. The former is plotted in blue, while the latter is in red.

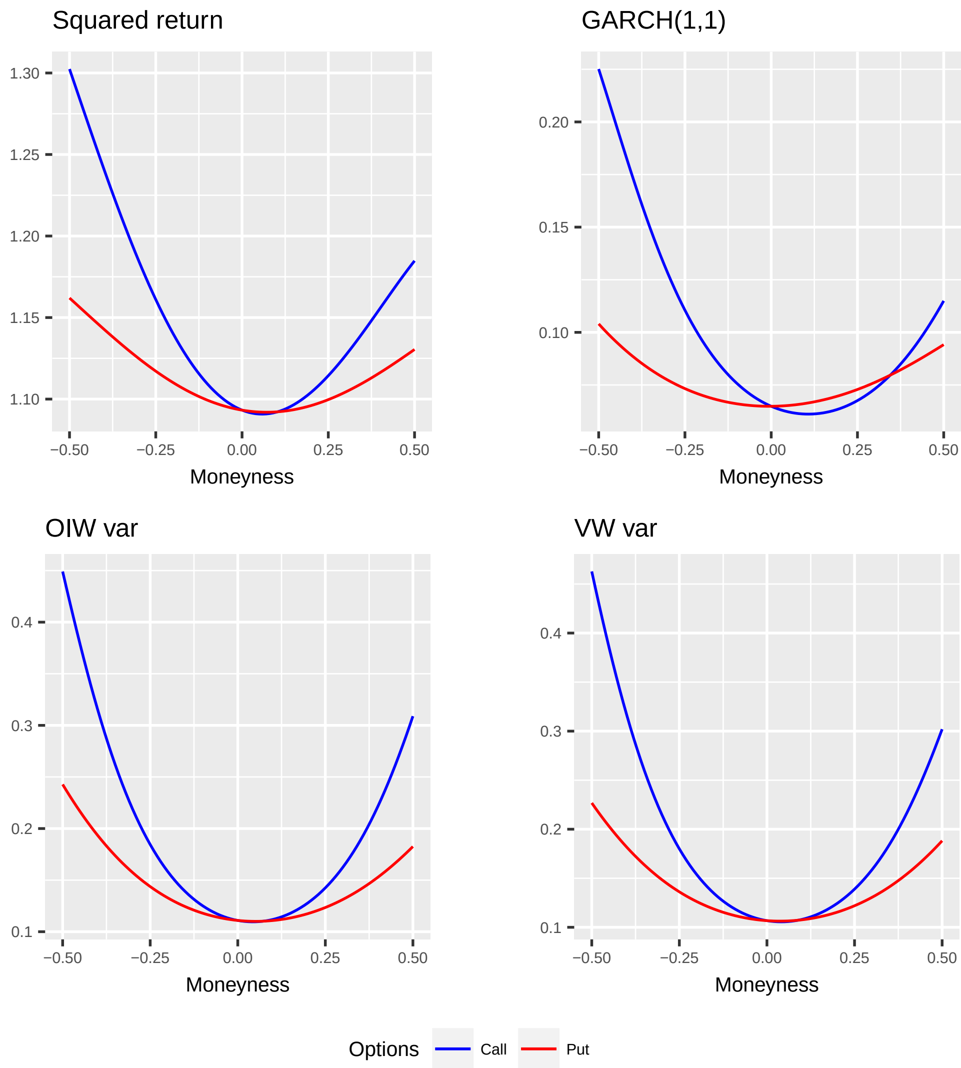


Table A3: Spillover Effect

This Table reports the outcome of the regression in Equation (1.2) where the dummy for expiration date has been replaced with two dummies: $D_{expiring\_options}$ is 1 for those stocks with at least one option expiring on a given expiration date, and $D_{no\_expiring\_options}$ for those stocks with no expiring options on a given expiration date. The table shows a significant positive spillover effect to those stocks with no option expiring. Each regression includes stock fixed effects and clusters the errors at the stock level.

	Dependent variable:			
	Log(σ^2)			
	Squared return	GARCH(1,1)	VW-IV	OIW-IV
	(1)	(2)	(3)	(4)
D(Friday)	-0.007^{***}	0.009^{***}	-0.042^{***}	-0.032^{***}
	(0.0004)	(0.001)	(0.001)	(0.001)
D(expiring options)	-0.002^{***}	-0.009^{***}	0.005	0.013^{***}
	(0.001)	(0.002)	(0.004)	(0.004)
D(no expiring options)	-0.001	0.037^{*}	0.214^{***}	0.183^{***}
	(0.002)	(0.020)	(0.022)	(0.022)
Stock FE	Y	Y	Y	Y
Observations	2,128,988	1,545,786	2,128,988	2,128,988
R^2	0.012	0.059	0.079	0.089
Adjusted R^2	0.012	0.058	0.079	0.089
Note:				*p<0.1; **p<0.05; ***p<0.01

Table A4: Regression of Volatility on Open-Interest-weighted Delta Hedging

This Table reports the outcome of the regression in Equation (1.2), to which a measure of investors' delta hedging is added. Delta hedging is computed as $\Delta F e^{r(T-t)}$, where F is the most liquidity future on the S&P500. Delta hedging is also interacted with the dummy for option expiration. A stock fixed effect is also included in the regression, and errors are clustered at the stock level.

	Dependent variable:			
	$\text{Log}(\sigma^2)$			
	Squared return	GARCH(1,1)	VW-IV	OIW-IV
	(1)	(2)	(3)	(4)
D(expiring options)	-0.002	0.021	0.049^{***}	0.058^{***}
	(0.005)	(0.020)	(0.016)	(0.016)
D(no expiring options)	0.001	-0.012	0.112^{***}	0.138^{***}
	(0.005)	(0.037)	(0.026)	(0.026)
Delta Hedging	0.020^{***}	0.193^{***}	0.340^{***}	0.355^{***}
	(0.001)	(0.003)	(0.002)	(0.002)
D(expiring options) ×	0.003	-0.016	-0.022^{**}	-0.018^{*}
Delta Hedging	(0.003)	(0.013)	(0.011)	(0.010)
Stock FE	Y	Y	Y	Y
Observations	891,946	798,331	891,946	891,946
R^2	0.021	0.084	0.114	0.126
Adjusted R^2	0.021	0.083	0.113	0.126
Note:				*p<0.1; **p<0.05; ***p<0.01

Table A5: Trading Strategy

This Table reports some summary statistics of the trading strategy discussed in section 1.5. Strategy 1 is a conservative strategy, where investors are more concerned about risk and choose options based on their liquidity. Strategy 2 is a more aggressive strategy, where investors only aim at maximizing the option premia. The column "TTE" indicates how many days before expiration the strategy is implemented. For each strategy, the expected return, the standard deviation, the annualized Sharpe Ratio and the annualized Information Ratio are reported. Panel A uses the whole sample of stocks, while in panel B we restrict the sample to stocks in the S&P500 only.

Panel A: All stocks

	Strategy 1				Strategy 2			
TTE	Ex ret	Std	SR	IR	Ex ret	Std	SR	IR
1	0.53	0.85	2.16	2.15	0.66	0.82	2.77	2.76
2	0.24	0.71	1.17	0.83	0.40	0.70	1.41	1.41
3	0.03	0.61	0.10	0.09	0.21	0.60	0.70	0.70
4	−0.06	0.59	−0.17	−0.17	0.12	0.54	0.37	0.36

Panel B: Options on Stocks in the S&P 500

	Strategy 1				Strategy 2			
TTE	Ex ret	Std	SR	IR	Ex ret	Std	SR	IR
1	0.63	1.09	1.99	1.98	0.74	1.08	2.39	2.38
2	0.32	0.94	0.84	0.83	0.49	0.93	1.30	1.30
3	0.09	0.65	0.27	0.26	0.30	0.63	0.96	0.96
4	−0.004	0.64	−0.01	−0.01	0.21	0.60	0.60	0.59

Table A6: Trading Strategy - Robustness Checks

This Table reports robustness checks of the trading strategy discussed in section 1.5. Strategy 1 is a conservative strategy, where investors are more concerned about risk and choose options based on their liquidity. Strategy 2 is a more aggressive strategy, where investors only aim at maximizing the option premia. The column "TTE" indicates how many days before expiration the strategy is implemented. For each strategy, the expected return, the standard deviation, the annualized Sharpe Ratio and the annualized Information Ratio are reported. In panel A, expiration dates are moved to a random day over a 4-week window around the real expiration date. In the so-called *placebo expiration dates*, Strategy 1 turns out to be unprofitable, while Strategy 2 generates a Sharpe Ratio very close to 0 even before transaction costs. Panel B and C use the whole sample of stock and the stocks in the S&P500 only, respectively, but restricts the time period considered to the second half of our sample (from 2001 to 2006). In fact, the derivative market increased significantly in size, hence liquidity, in the early 2000s. Additionally, in the first half of the sample (from 1996 to 2000) financial markets were experiencing the Tech bubble, which may affect the results reported in Table A5.

Panel A: Placebo expiration dates

	Strategy 1				Strategy 2			
TTE	Ex ret	Std	SR	IR	Ex ret	Std	SR	IR
1	−0.04	0.74	−0.18	−0.18	0.07	0.69	0.33	0.33
2	−0.07	0.65	−0.24	−0.25	0.05	0.65	0.18	0.18
3	−0.08	0.66	−0.24	−0.24	0.02	0.65	0.07	0.06
4	−0.12	2.92	−0.07	−0.07	−0.02	2.93	−0.01	−0.01

Panel B: All Stocks - second half

	Strategy 1				Strategy 2			
TTE	Ex ret	Std	SR	IR	Ex ret	Std	SR	IR
1	0.56	0.76	2.52	2.51	0.67	0.75	3.09	3.08
2	0.27	0.67	0.99	0.98	0.43	0.68	1.55	1.54
3	0.06	0.53	0.24	0.23	0.23	0.54	0.86	0.85
4	−0.02	0.54	−0.06	−0.07	0.14	0.53	0.45	0.45

Panel C: Options on S&P 500 only - second half

	Strategy 1				Strategy 2			
TTE	Ex ret	Std	SR	IR	Ex ret	Std	SR	IR
1	0.64	0.95	2.33	2.32	0.74	0.94	2.75	2.74
2	0.35	0.92	0.92	0.92	0.53	0.94	1.37	1.36
3	0.12	0.57	0.44	0.43	0.33	0.57	1.16	1.16
4	0.03	0.61	0.09	0.09	0.23	0.58	0.69	0.68

Table A7: Robustness Check: controlling for VIX

This table reports the outcome of regression in Equation (1.2), with the inclusion of a control for the market volatility, measured by the VIX index. Specifically, three dummies are added to the regression: one for days of low VIX (below 5th percentile), one for days with medium VIX (between 45th and 55th percentiles) and one for days with high VIX (above 95th percentile). Low (High) VIX dummy coefficients are negative (positive), showing that all four volatility measures are significantly lower (higher) than the average on days of low (high) VIX. Medium VIX coefficients are positive but very low in magnitude. The negative effect on realized volatility and positive effect on implied volatility on option expiration date persist. Each regression includes a stock fixed effect, and errors are clustered at the stock level.

	Dependent variable:			
	$\text{Log}(\sigma^2)$			
	Squared return	GARCH(1,1)	VW-IV	OIW-IV
	(1)	(2)	(3)	(4)
D(expiring options)	-0.002^{***}	-0.003^{*}	0.008^{***}	0.006^{***}
	(0.001)	(0.002)	(0.004)	(0.004)
D(no exp. options)	-0.002	0.036^{*}	0.207^{***}	0.175^{***}
	(0.002)	(0.019)	(0.022)	(0.022)
Low VIX	-0.038^{***}	-0.508^{***}	-0.343^{***}	-0.370^{***}
	(0.001)	(0.013)	(0.013)	(0.013)
Medium VIX	-0.001	0.022^{***}	0.018^{***}	0.023^{***}
	(0.001)	(0.005)	(0.005)	(0.004)
High VIX	0.097^{***}	0.601^{***}	0.443^{***}	0.415^{***}
	(0.002)	(0.011)	(0.010)	(0.010)
Stock FE	Y	Y	Y	Y
Observations	2,217,693	1,544,970	2,217,693	2,217,693
R^2	0.027	0.122	0.112	0.123
Adjusted R^2	0.027	0.122	0.111	0.123
Note:				*p<0.1; **p<0.05; ***p<0.01

Table A8: Robustness Check: controlling for Net Buying Pressure

This table reports the outcome of regression in Equation (1.2), with the inclusion of a control for the net buying pressure. Net buying pressure is defined as the difference between buyer-motivated and seller-motivated trades. Buyer-motivated trades are those that occur at a price above the daily most frequent bid-ask midpoint. Seller-motivated trades are those that occur at a price below the daily most frequent bid-ask midpoint. As in Bollen & Whaley (2004), net buying pressure positively correlates with implied volatility. Each regression includes a stock fixed effect, and errors are clustered at the stock level.

	Dependent variable:			
	$\text{Log}(\sigma^2)$			
	Squared return	GARCH(1,1)	VW-IV	OIW-IV
	(1)	(2)	(3)	(4)
D(expiring options)	−0.002	−0.005**	0.010*	0.015
	(0.001)	(0.002)	(0.005)	(0.005)
D(no exp. options)	−0.002	0.034	0.191***	0.172***
	(0.002)	(0.023)	(0.034)	(0.033)
Net buying pressure	−0.002***	0.001	0.002***	0.002**
	(0.0004)	(0.001)	(0.001)	(0.001)
Stock FE	Y	Y	Y	Y
Observations	1,111,175	857,219	1,111,175	1,111,175
R^2	0.014	0.059	0.088	0.100
Adjusted R^2	0.013	0.059	0.087	0.099
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure A2: Implied Volatility around Expiration Dates

These figures show how each implied volatility measure evolves as we approach the option expiration date, which is set at time 0. For each plot, confidence bands at 0.05% and 99.5% are included, obtained by running a Monte Carlo simulation with 1,000 draws from a Normal-Inverse-Wishart distribution. The y-axis reports the logarithm of implied volatility, hence the negative range.

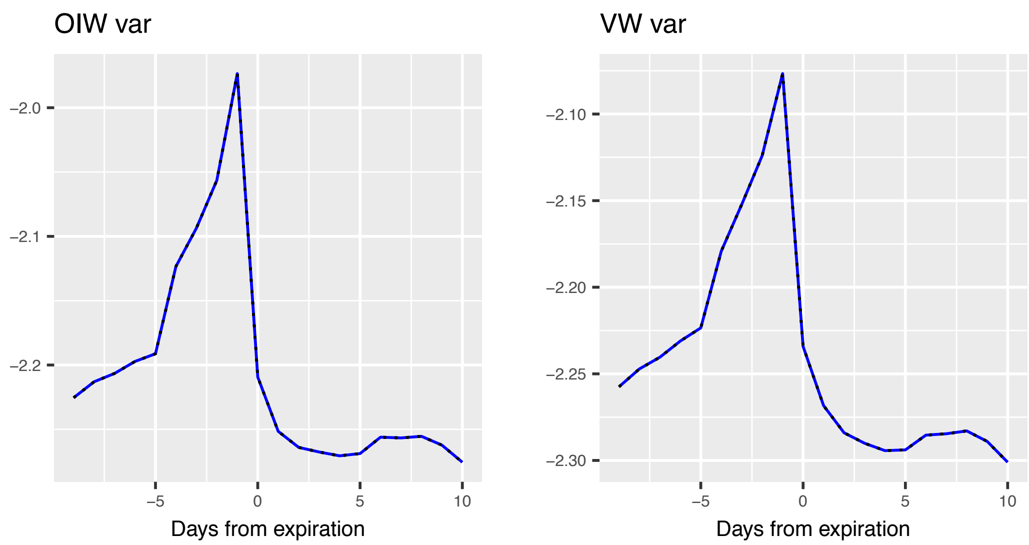


Figure A3: Boxplot of Distance of Stock Price from Strike around Expiration

Boxplots of percentage of stocks closing closer than \$0.5, \$0.2 and \$0.1 to the closest strike price over a 20-day window around expiration. On expiration dates, the first quartile is above the median value of any other day. This suggests that there is evidence of pinning in our sample.

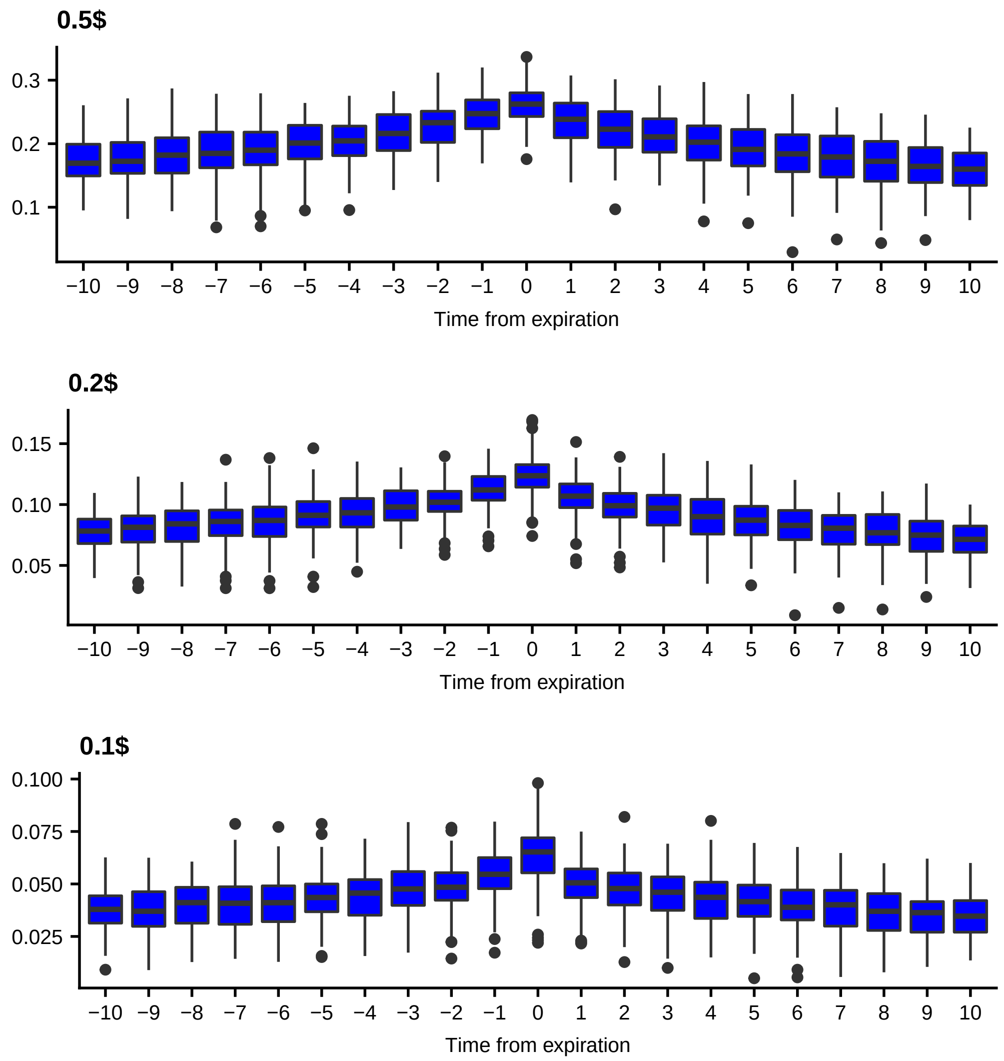


Figure A4: Payoff Structure of a Straddle

A straddle is a combination of options that consists in shorting both a call and a put on a given underlying asset. The red dashed line is the payoff of a short call, while the blue dashed line is the payoff of a short put. Both of them are shifted upward by the premia received from selling the options. The black solid line is the net payoff from the straddle. If the realized volatility turns out to be smaller than the implied volatility (as we expect on option expiration dates), the stock price at expiration will be in a neighbourhood of the strike price S_0, and the net payoff will be positive (green area). If instead realized volatility exceeds implied volatility, this strategy generates a loss. If the stock price drops, the short put is exercised, while if the stock price increases, the short call is exercised. In both cases, the option writer suffers a loss (red area).

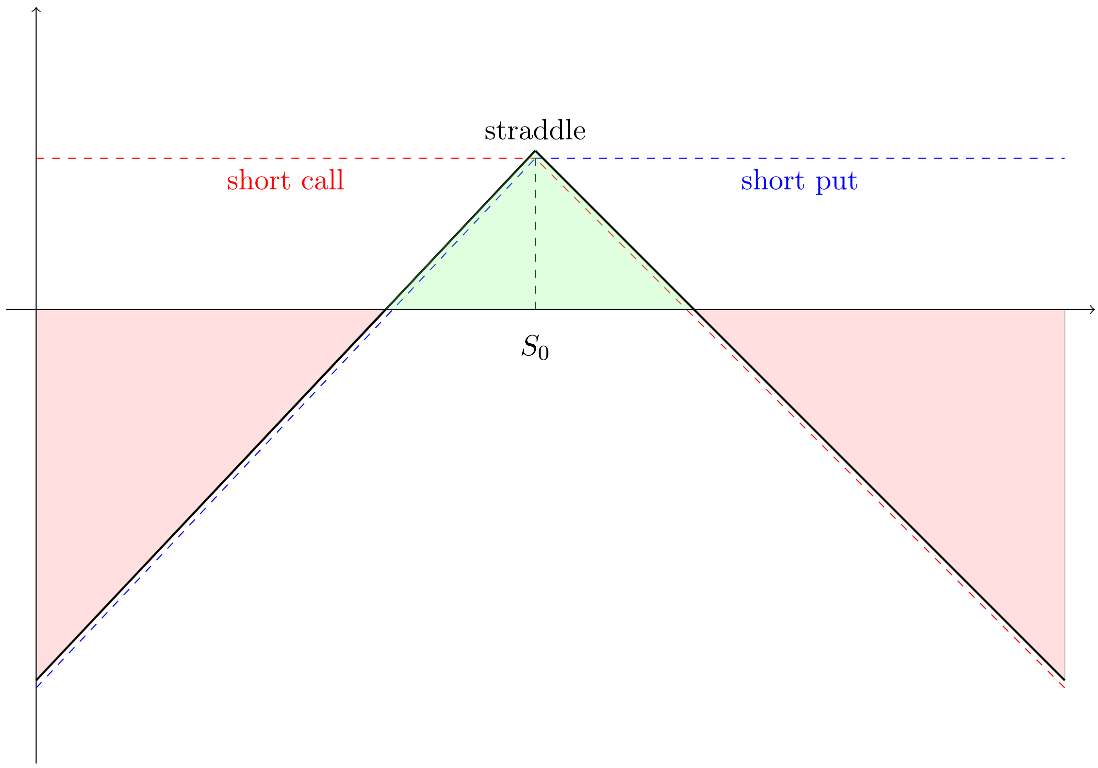


Figure A5: Volatility Smile of Call Options (controlling for VIX)

For each volatility measure, this Figure plots the volatility of call options as a function of moneyness. The blue line shows a benchmark case, derived from the six-order polynomial of regression in Equation (1.2), whose output is shown in Table A2. The red line plots the same six-order polynomial from a regression which controls for the VIX index. Specifically, three dummies are added to the regression: one for days of low VIX (below 5th percentile), medium VIX (between 45th and 55th percentiles) and high VIX (above 95th percentile). No significant difference can be observed, especially for at-the-money options.

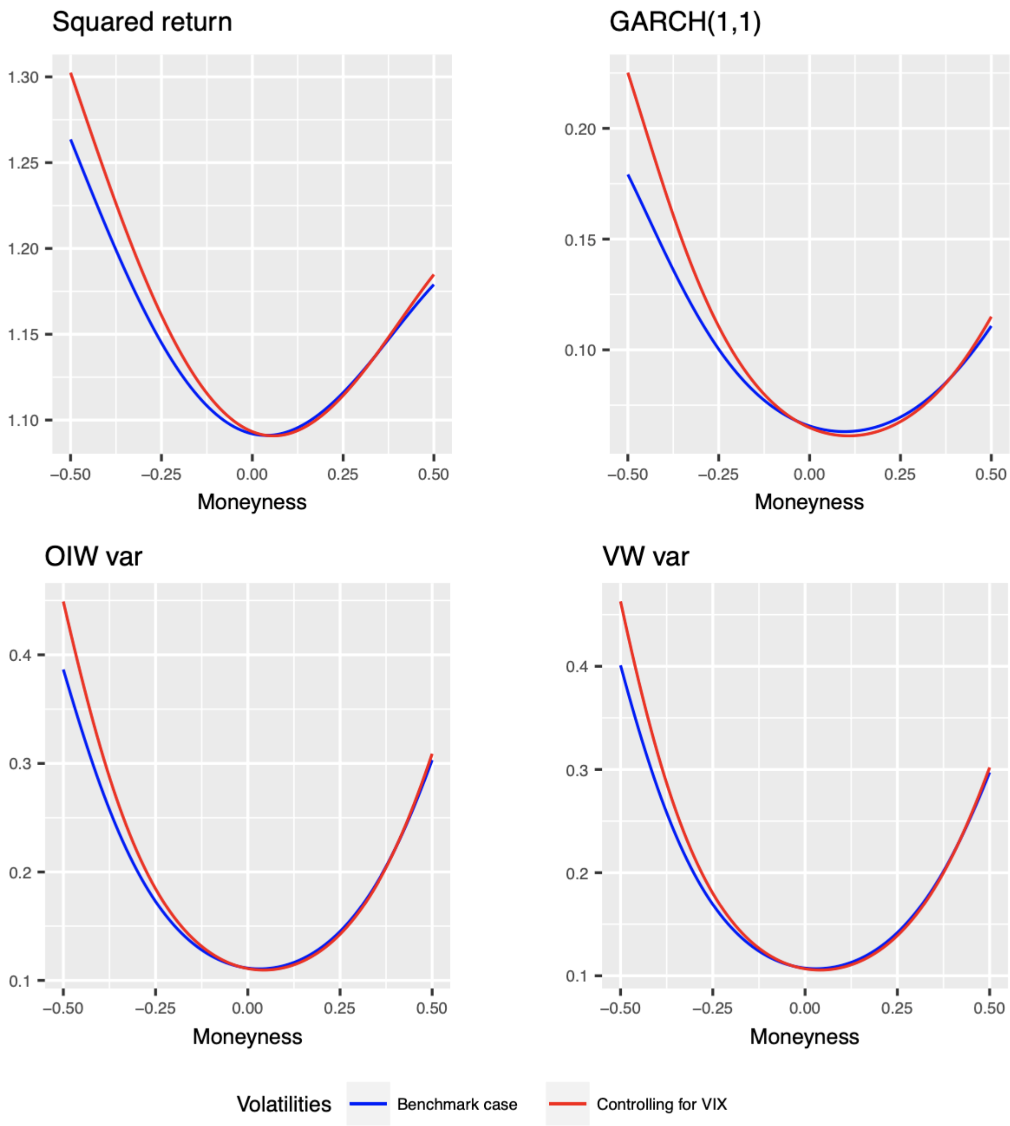
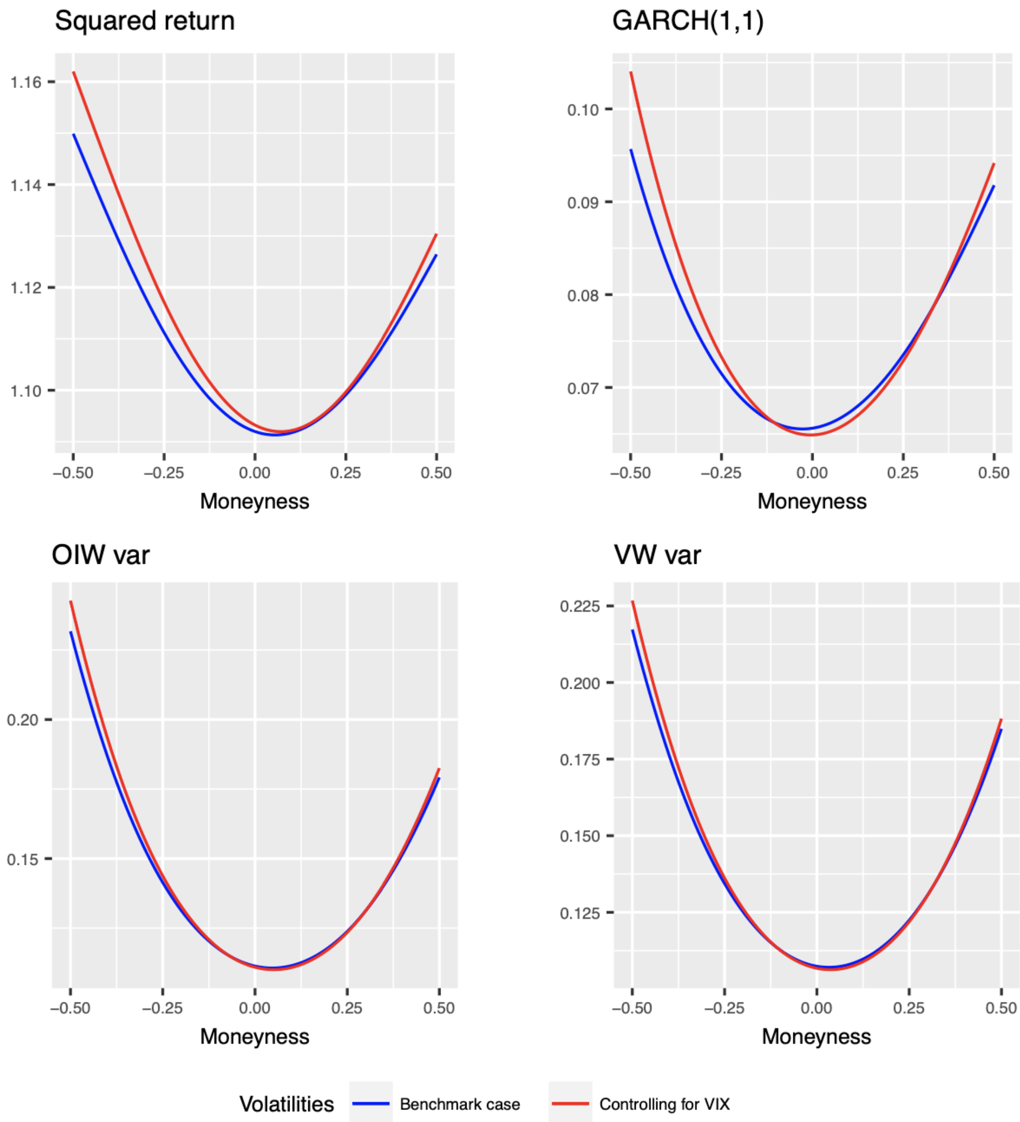


Figure A6: Volatility Smile of Put Options (controlling for VIX)

For each volatility measure, this Figure plots the volatility of put options as a function of moneyness. The blue line shows a benchmark case, derived from the six-order polynomial of regression in Equation (1.2), whose output is shown in Table A2. The red line plots the same six-order polynomial from a regression which controls for the VIX index. Specifically, three dummies are added in the regression: one for days of low VIX (below 5th percentile), medium VIX (between 45th and 55th percentiles) and high VIX (above 95th percentile). No significant difference can be observed, especially for at-the-money options.



B. Appendix to Mutual Funds' Behaviour during Financial Bubbles

Figure B1: Timing of a Bubble Episode

Time 0 is defined as the first month in which a sector experiences a 100% or more raw and excess return. A bubble episode runs from $t = -24$ to $t + 30$. The peak is located after time 0, on average around $t = 8$. The period from $t = -24$ to *Peak* is called *Run-up* while the period from *Peak* to $t = 30$ is called *After Run-up*

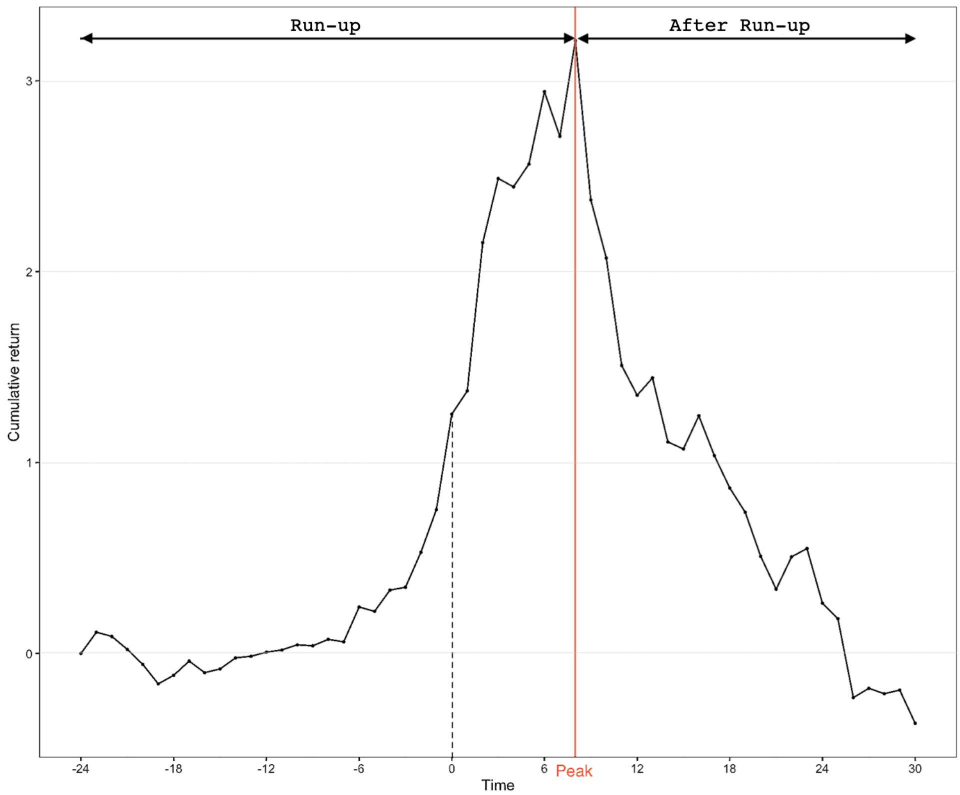


Table B1: Bubble Episodes

This Table displays the summary statistics of the main bubble episodes for the period 1980-2018. Panel A reports those episodes which eventually crash. For each of them, column 2 contains the first quarter after a 100% raw and excess return, and is the time 0 for that episode. Column 3 shows the number of firms in each sector. Columns 4 and 5 display the 2-year raw and excess return (not annualized) while column 6 contains the smallest return (largest loss) after the peak. Finally, columns 7 and 8 report the number of months from time 0 to when the peak is reached and the crash occurs, respectively. Panel B reports the same feature for "fundamentally driven" run-ups, i.e. those run-ups which are not followed by a crash. In these cases, no peak can be identified, nor a crash is observed.

Sector	Time	Nr. firms	2yr Ret	2yr Ex Ret	Crash Ret	Peak	Crash
Panel A - Episodes with Crash							
Chips	1999/12	347	2.06	1.16	0.42	6	9
Hardware	1999/03	171	2.28	1.01	0.26	6	9
Software	1999/03	640	2.38	1.20	0.56	12	15
Finance	2000/03	150	1.77	1.00	0.55	3	6
Smoke	2001/12	3	1.14	1.45	0.57	3	6
Other	2001/03	33	1.03	1.10	0.53	3	6
Real Estate	2006/03	27	1.44	1.05	0.56	12	15
Mines	2006/06	16	1.73	1.20	0.40	24	27
Steel	2007/06	39	1.86	1.16	0.30	9	12
Coal	2008/06	13	1.40	1.28	0.21	0	3
Mean		144	1.71	1.16	0.44	7.5	10.5
Panel B - Episodes with No Crash							
Gold	1987/12	38	1.40	1.04			
Mines	1988/03	21	1.30	1.09			
Medical Equip.	1991/09	195	1.43	1.11			
Toys	1992/06	49	1.38	1.00			
Chips	1993/09	309	1.59	1.01			
Entertainment	1993/09	112	1.74	1.12			
Telecommunication	1999/09	201	1.79	1.02			
Healthcare	2000/12	92	1.48	1.07			
Laboratory Equip.	2000/03	119	2.14	1.25			
Real Estate	2001/12	36	1.12	1.39			
Construction	2001/03	63	1.09	1.17			
Electrical Equip.	2001/03	82	1.36	1.17			
Wholesales	2001/06	189	1.03	1.04			
Clothing	2002/03	68	1.10	1.56			
Personal Services	2002/03	59	1.13	1.73			
Automotive	2010/09	53	1.17	1.28			
Entertainment	2010/09	47	1.13	1.33			
Real Estate	2010/09	25	1.19	1.38			
Entertainment	2018/06	49	1.79	1.02			
Mean		72	1.39	1.25			

Figure B2: Evolution of Excess Holdings during a Bubble Episode

This Figure plots the δ coefficients from Equation (2.3). The solid line plots the point estimates, while the dashed lines represent the 95% confidence interval. All bubble episodes are aligned on the peak, which is indicated by the red vertical dashed line. Starting from 5 quarters before the peak, the δ coefficients are positive and significant, showing a gradual over-investment of funds. Mutual funds withdraw money at the peak, which is the first time that δ becomes negative and significant. From 6 months after the peak, they start investing again.

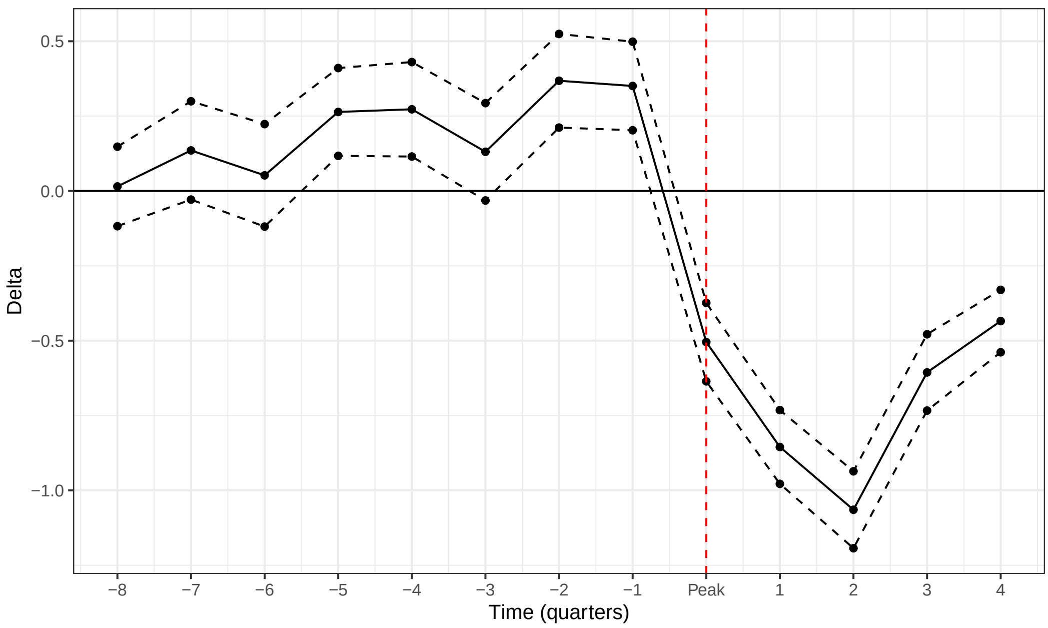
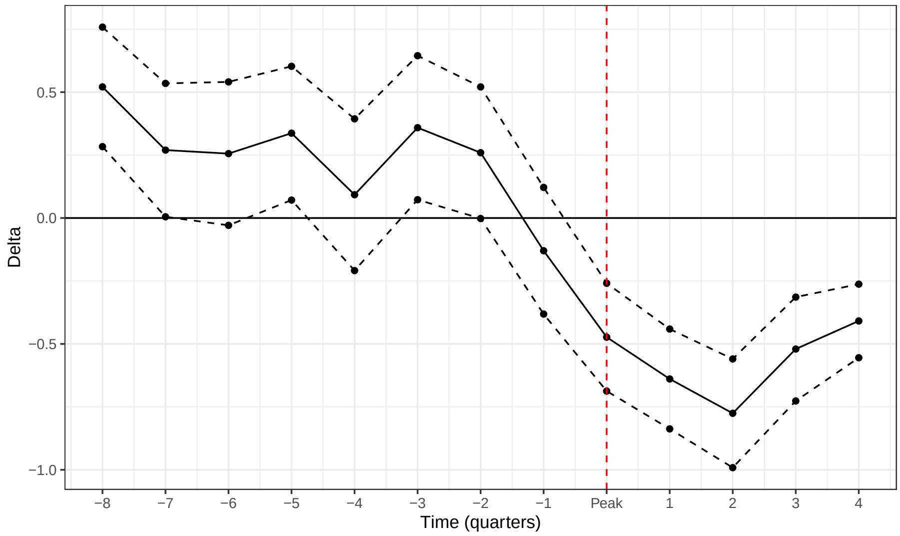
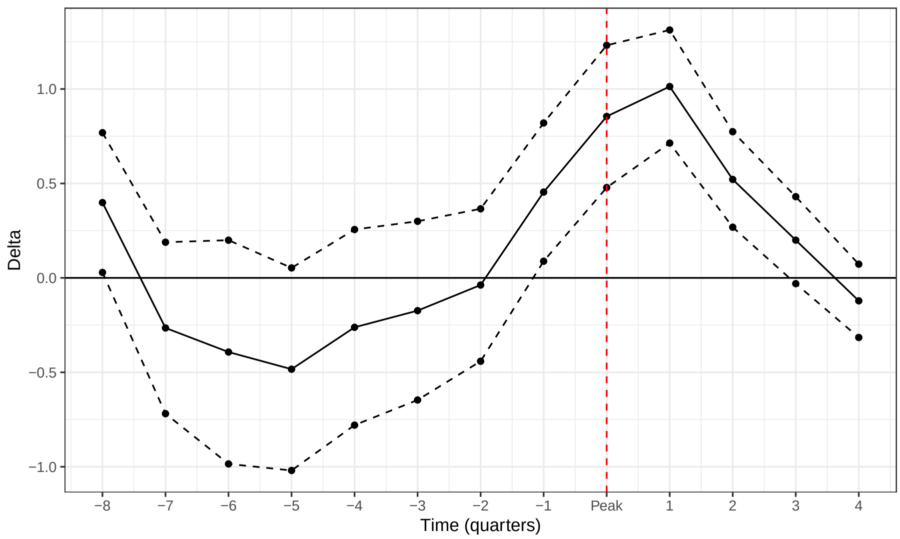


Figure B3: Evolution of Excess Holdings during a Bubble Episode

This Figure plots the δ coefficients from Equation (2.3). In the top panel, the sample is restricted to growth funds only, whose holdings are compared to an index of growth stocks. Similarly, in the bottom panel, the sample is restricted to value funds, whose holdings are compared to an index of value stocks. The solid lines plot the point estimates, while the dashed lines represent the 95% confidence intervals. All bubble episodes are aligned on the peak, which is indicated by the red vertical dashed lines. The two plots show a very different behaviour between growth and value funds, with growth funds proactively exploit the price run-up.



(a) Growth



(b) Value

Figure B4: Evolution of Excess Holdings during a Bubble Episode.

This Figure plots the δ coefficients of the regression displayed in Equation (2.9). The solid line plots the point estimates, while the dashed lines represent the 95% confidence interval. All bubble episodes are aligned on the quarter the run-up is first identified, which corresponds to time 0. The average peak (for episodes that crash) is indicated by the red vertical dashed line. No statistically different behaviour is observed during the run-up up to one quarter before the peak.

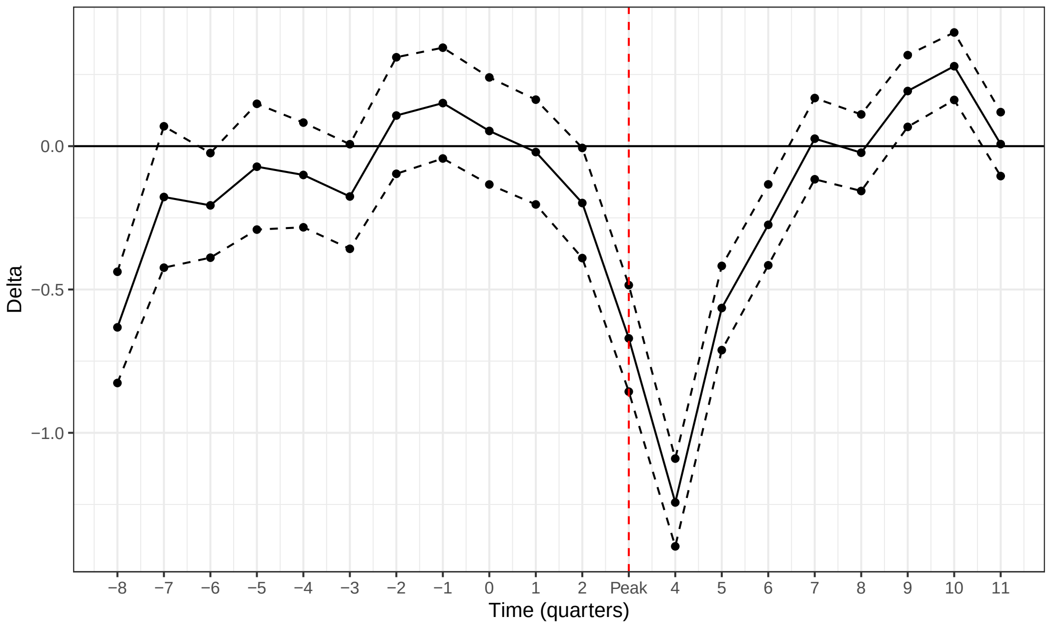


Table B2: Excess Holdings

This Table displays the outcome of the regression of excess holdings on the dummy *Run-up*. In column (1), excess holdings are computed with respect to the whole universe of stocks in CRSP. In column (2) the sample is restricted to stocks whose price is larger than 5 dollars. In columns (3) and (4) the sample is restricted to growth and value funds and their excess holdings are then constructed in comparison to an index made of growth and value stocks only, respectively. All regressions include time, sector and fund fixed effects. Errors are clustered at the fund level.

	Dependent variable:			
	Excess Holdings			
	CRSP	Price	Growth	Value
	(1)	(2)	(3)	(4)
Run-up	0.117***	0.106***	0.379***	0.417***
	(0.034)	(0.034)	(0.052)	(0.076)
Time FE	Y	Y	Y	Y
Fund FE	Y	Y	Y	Y
Sector FE	Y	Y	Y	Y
R^2	0.308	0.308	0.320	0.458
Adjusted R^2	0.307	0.307	0.319	0.457
Note:				*p<0.1; **p<0.05; ***p<0.01

Table B3: Stock Picking Ability (holdings)

This Table shows the output of the regression displayed in Equation (2.4). In the first two columns the whole sample is used, while in the last two columns the sample is restricted to active funds only. For both samples, two regressions are run: the first uses the 3-month $\hat{\alpha}$ as dependent variable, while the second uses the 6-month $\hat{\alpha}$. All regressions include time, sector and fund fixed effects. Errors are clustered at the fund level.

	Dependent variable:			
	All Funds		Active Funds	
	3m	6m	3m	6m
	(1)	(2)	(3)	(4)
Excess Holdings	0.003*	0.002	0.002*	0.005
	(0.001)	(0.004)	(0.001)	(0.003)
Run-up × Excess Holdings	0.184***	0.339***	0.190***	0.372***
	(0.015)	(0.029)	(0.015)	(0.030)
Time FE	Y	Y	Y	Y
Sector FE	Y	Y	Y	Y
Fund FE	Y	Y	Y	Y
Observations	2,752,690	2,752,690	2,358,582	2,358,582
R^2	0.197	0.076	0.202	0.081
Note:				*p<0.1; **p<0.05; ***p<0.01

Table B4: Stock Picking Ability (trades)

This Table shows the output of the regression displayed in Equation (2.5). In the first two columns the whole sample is used, while in the last two columns the sample is restricted to active funds only. For both samples, two regressions are run: the first uses the 3-month $\hat{\alpha}$ as dependent variable, while the second uses the 6-month $\hat{\alpha}$. All regressions include time, sector and fund fixed effects. Errors are clustered at the fund level.

	Dependent variable:			
	All Funds		Active Funds	
	3m	6m	3m	6m
	(1)	(2)	(3)	(4)
Trades	-0.006^{***}	-0.010^{***}	-0.006^{***}	-0.007^{*}
	(0.002)	(0.003)	(0.002)	(0.004)
Run-up × Trades	0.040^{***}	0.041^{***}	0.033^{***}	0.028^{*}
	(0.007)	(0.013)	(0.009)	(0.016)
Time FE	Y	Y	Y	Y
Sector FE	Y	Y	Y	Y
Fund FE	Y	Y	Y	Y
Observations	564,097	564,097	437,305	437,305
R^2	0.174	0.073	0.185	0.074
Note:				*p<0.1; **p<0.05; ***p<0.01

Table B5: Mutual Fund Excess Returns

This Table displays the regressions of mutual fund excess return on the Fama French factors and the dummy *Run-up*. Column (1) shows a CAPM regression, column (2) also includes the SMB and HML factors from Fama & French (1993), while columns (3) and (4) add RMW (Robust Minus Weak) and CMA (Conservative Minus Aggressive) factors from Fama & French (2015). All regressions include fund fixed effects. The coefficients on the dummy *Run-up* are all positive and statistically significant, showing that mutual funds earn abnormal profits during financial bubbles.

	Dependent variable:			
	Mutual Fund Excess Returns			
	(1)	(2)	(3)	(4)
Run-up	0.013***	0.013***	0.013***	0.015***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Excess Market Return	1.088***	1.023***	1.024***	1.034***
	(0.001)	(0.001)	(0.001)	(0.001)
SMB		0.324***	0.325***	0.320***
		(0.002)	(0.002)	(0.002)
HML		−0.052***	−0.054***	−0.116***
		(0.002)	(0.002)	(0.002)
RMW			0.006**	0.001
			(0.002)	(0.002)
CMA				0.132***
				(0.004)
Fund FE	Y	Y	Y	Y
R^2	0.533	0.543	0.543	0.544
Adjusted R^2	0.520	0.530	0.530	0.531
Note:				*p<0.1; **p<0.05; ***p<0.01

Table B6: Excess Holdings (value-weighted)

This Table reports the outcome of the regression in Equation (2.7). Mutual funds are aggregated on a value-weighted basis and excess holdings are regressed on the dummy *Run-up* and some industry characteristics, namely size and past return. The regression includes time and sector fixed effects.

	Dependent variable:		
	Excess holdings		
	CRSP	Growth	Value
Run-up	−0.568	−0.265	0.187
	(1.208)	(1.337)	(1.733)
Time FE	Y	Y	Y
Sector FE	Y	Y	Y
R^2	0.428	0.483	0.585
Adjusted R^2	0.360	0.415	0.528
Note:			*p<0.1; **p<0.05; ***p<0.01

Table B7: Contemporaneous Correlation between Excess Returns and Trades

This Table shows the output of the regression displayed in Equation (2.8). In the first two columns the whole sample is used, while in the last two columns the sample is restricted to active funds only. For both samples, two regressions are run: the first uses the 3-month $\hat{\alpha}$ as dependent variable, while the second uses the 6-month $\hat{\alpha}$. The variable *Trades* and the dependent variable $\hat{\alpha}$ are contemporaneous, i.e. they are calculated over the same time period. All regressions include time, sector and fund fixed effects. Errors are clustered at the fund level.

	Dependent variable:			
	All Funds		Active Funds	
	3m	6m	3m	6m
	(1)	(2)	(3)	(4)
Trades	−0.0004	0.030***	−0.001	0.040***
	(0.001)	(0.007)	(0.002)	(0.009)
Run-up × Trades	0.003	−0.035**	−0.006	−0.070***
	(0.008)	(0.018)	(0.010)	(0.021)
Time FE	Y	Y	Y	Y
Sector FE	Y	Y	Y	Y
Fund FE	Y	Y	Y	Y
Observations	486,089	486,089	367,677	367,677
R^2	0.173	0.074	0.188	0.077
Note:				*p<0.1; **p<0.05; ***p<0.01

Table B8: Return Predictability of Mutual Fund Holdings

This Table reports the coefficient β, its standard error and the R-squared of regression in Equation (2.6). Each row displays a different holding characteristic, namely the excess holdings on a value-weighted index of all the stocks in CRSP, the difference in holdings between growth and value funds (GMV), acceleration and turnover. Acceleration is defined as the change in number of shares held over the last two quarters over the change in the number of shares held in the last 4 quarters $\left(\frac{\text{Shares}_{i,t}-\text{Shares}_{i,t-2}}{\text{Shares}_{i,t}-\text{Shares}_{i,t-4}}\right)$. Turnover is the percentage change in shares held $\left(\frac{\Delta\text{shares}_{i,t}}{\text{Shares}_{i,t-1}}\right)$. Regressions are run for three different time horizons, from 3 months, to 6 months, to 1 year. They all include sector and time fixed effects, which are not reported.

	Dependent variable:								
	Excess Market Return								
	3 months			6 months			1 year		
	β	s.e.	R^2	β	s.e.	R^2	β	s.e.	R^2
Excess Holdings	**2.198**	(0.662)	0.338	**4.211**	(1.113)	0.407	**5.764**	(1.529)	0.435
GMV	**0.012**	(0.004)	0.390	**0.020**	(0.009)	0.461	**0.043**	(0.015)	0.502
Acceleration	**0.093**	(0.044)	0.344	0.092	(0.064)	0.456	**0.261**	(0.071)	0.528
Turnover	**0.014**	(0.006)	0.324	**0.014**	(0.005)	0.383	**0.028**	(0.014)	0.420

C. Appendix to On the Time-Series Persistence of Mutual Fund Fees

Table C1: Sector Classification

This Table describes how to link the CRSP sector classification to the Fama-French sector classification. In fact, mutual funds are associated to a CRSP sector based on they style. While stocks, whose returns are used to define bubbles, are associated to a Fama-French sector based on their SIC codes. As CRSP classification is less granular, and to each CRSP sector corresponds one or more Fama-French sectors. This mapping allows to identify which funds have an investing style which focuses on a sector that experiences a financial bubble.

CRSP	**Fama-French**
Commodity	Agriculture, Rubber and Plastic Products
Consumer Goods	Apparel, Beer & Liquor, Business Supplies, Candy & Soda, Consumer Goods, Food Products, Printing and Publishing, Recreation, Retail, Tobacco Products, Wholesale
Consumer Services	Business Services, Entertainment, Personal Services, Restaurants, Hotels and Motels, Transportation
Financial	Banking, Insurance, Trading
Gold	Precious Metals
Healthcare	Healthcare
Industrial	Aircraft, Automobiles/Trucks, Chemicals, Defense, Electrical Equipment, Laboratory Equipment, Machinery, Medical Equipment, Pharmaceutical Products, Shipbuilding and Railroad Equipment
Materials	Building Materials, Construction, Construction Materials, Fabricated Products, Shipping Containers, Textiles
Natural Resources	Coal, Non-Metallic and Industrial Metal Mining, Petroleum and Natural Gas, Steel
Real Estate	Real Estate
Technology	Electronic Equipment, Hardware, Software
Telecommunication	Communication
Utility	Utility

Table C2: Summary Statistics

For each type of fee, this Table reports some summary statistics, namely the minimum, the maximum, the average, the quartiles and the number of non-missing values (in thousands). Management fees may be negative due to waivers and reimbursements. Not all funds charge load and rear fees, hence their minimum values are 0. 12b-1 fees are associated to marketing costs and cannot exceed 1%. Expense ratio is calculated as the percentage of total investment that shareholders pay for the fund's operating expenses. All numbers, except the number of observations, are in percent.

	Min	1st Q	Median	Mean	3rd Q	Max	N (th)
Mgmt Fees	-1.569	0.200	0.524	0.342	0.759	1.243	1,149
Front Load	0	0	2.500	2.052	3	5.750	359
Rear Load	0	0	0.500	0.723	1	2.500	648
12b-1 Fees	0.025	0.250	0.500	0.557	1	1	707
Expense Ratio	0.080	0.670	1.100	1.153	1.580	2.500	1,251

Table C3: Fees during Bubble Episodes

This Table reports the regression outcome of Equation (3.1). Each column corresponds to a different regression in which only the dependent variable changes. Only the independent variable of interest is shown. *Run-up* is a dummy that takes the value 1 if a sector a fund operates in is experiencing a price run-up. Each regression also includes a year fixed effect and clusters errors at the fund level.

	Dependent variable:				
	Management Fees	12b-1	Front Load	Rear Load	Expense Ratio
	(1)	(2)	(3)	(4)	(5)
Run-up	-0.024^{*}	0.012^{***}	0.037^{***}	0.060^{***}	0.013^{***}
	(0.014)	(0.001)	(0.007)	(0.003)	(0.002)
Year FE	Y	Y	Y	Y	Y
Observations	1,127,056	641,698	312,920	564,923	1,126,814
R^2	0.001	0.044	0.096	0.155	0.069
Adjusted R^2	0.001	0.044	0.096	0.155	0.069
Note:					*p<0.1; **p<0.05; ***p<0.01

Table C4: Fees during Bubble Episodes (controlling for competition)

This Table reports the regression outcome of Equation (3.1), with the addition of a control variable for competition in the mutual fund industry. Each column corresponds to a different regression in which only the dependent variable changes. Only the independent variable of interest is shown. *Run-up* is a dummy that takes the value 1 if a sector a fund operates in is experiencing a price run-up. *MVO* measures competition and is computed as the average percent holding that each fund shares with any other fund. Each regression also includes a year fixed effect and clusters errors at the fund level.

	Dependent variable:				
	Management Fees	12b-1	Front Load	Rear Load	Expense Ratio
	(1)	(2)	(3)	(4)	(5)
Run-up	0.054**	0.055***	−0.126	0.042	0.097***
	(0.021)	(0.020)	(0.097)	(0.047)	(0.033)
MVO	−0.070	−0.409***	1.133***	−0.153	−0.717***
	(0.088)	(0.068)	(0.404)	(0.167)	(0.113)
Year × Sector FE	Y	Y	Y	Y	Y
Observations	5,595	3,811	1,637	3,296	5,951
R^2	0.086	0.184	0.348	0.229	0.308
Adjusted R^2	0.053	0.142	0.277	0.185	0.284
Note:					*p<0.1; **p<0.05; ***p<0.01

Table C5: Regression of Market Share on Fees

This Table reports the regression outcome of Equation (3.2). The dependent variable is the logarithm of market share. Only the independent variables of interest are included: *Expense ratio* is the percentage of total investment that investors pay for the fund's operating expenses, while *Bubble* is a dummy that takes a value of 1 if the sector the fund operates in is experiencing a financial bubble. The regression also includes fund and sector fixed effects and clusters errors at the fund level.

	Dependent variable:
	Log(Market Share)
Expense Ratio	−0.551***
	(0.135)
Expense Ratio × Bubble	0.359***
	(0.103)
Fund FE	Y
Sector FE	Y
R^2	0.857
Adjusted R^2	0.855
Note:	*p<0.1; **p<0.05; ***p<0.01

Table C6: IV Regression of Change in Market Share

This Table shows the IV regression of Equation (3.5). The dependent variable is the logarithm of market share. Due to simultaneity problem typical of any demand function estimation, an instrument for the fees is needed. To instrument fees charged by a given fund, the average fee charged by funds belonging to a different classification group is used. The cross-sectional long-term downward trend of fees ensures that the instrumented variable and the instrument are correlated. The exclusion restriction is satisfied as the instrument is computed using funds which are not direct competitors. The first stage regression includes sector and year fixed effect, but not a fund fixed effect to ensure that the fitted values are sufficiently different from the instrumented values. Fund, year and sector fixed effects are included in the second stage. Two IV estimation are reported: *No Bubble* excludes sectors which are experiencing a financial bubble, while *Bubble* restricts the sample to those funds that operate in a bubble sector.

	Dependent variable:			
	Log(Mkt Share)			
	No Bubble		Bubble	
	I Stage	II Stage	I Stage	II Stage
Instrument	0.928***		1.249***	
	(0.146)		(0.333)	
Expense Ratio		−2.896***		−0.233
		(0.539)		(0.397)
Fund FE	N	Y	N	Y
Sector FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
R^2	0.850	0.972	0.964	0.973
Adjusted R^2	0.848	0.971	0.961	0.971
Note:				*p<0.1; **p<0.05; ***p<0.01

Table C7: IV Regression of Market Share

This Table shows the IV regression of Equation (3.5). The dependent variable is the logarithm of change in market share. Due to simultaneity problem typical of any demand function estimation, an instrument for the fees is needed. To instrument fees charged by a given fund, the average fee charged by funds belonging to a different classification group is used. The cross-sectional long-term downward trend of fees ensures that the instrumented variable and the instrument are correlated. The exclusion restriction is satisfied as the instrument is computed using funds which are not direct competitors. The first stage regression includes sector and year fixed effect, but not a fund fixed effect to ensure that the fitted values are sufficiently different from the instrumented values. Fund, year and sector fixed effects are included in the second stage. Two IV estimation are reported: *No Bubble* excludes sectors which are experiencing a financial bubble, while *Bubble* restricts the sample to those funds that operate in a bubble sector.

	Dependent variable:			
	Log(ΔMkt Share)			
	No Bubble		Bubble	
	I Stage	II Stage	I Stage	II Stage
Instrument	0.930***		1.250***	
	(0.146)		(0.336)	
Expense Ratio		−2.894***		−0.579
		(0.538)		(1.008)
Fund FE	N	Y	N	Y
Sector FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
R^2	0.850	0.972	0.964	0.973
Adjusted R^2	0.848	0.971	0.961	0.971
Note:				*p<0.1; **p<0.05; ***p<0.01

Table C8: GMM Estimation

This Table reports the GMM estimation of Equation (3.9). The first three parameters, i.e. γ_1, γ_2 and γ_3, are used to model the marginal cost. The remaining two variables, i.e. θ_1 and θ_2, measure the competition as a function of the Herfindal index and are used to estimate the mark-up. Two GMM estimations are reported: *All Data* uses the whole sample, while *Bubble* restricts the sample to those funds that operate in a sector that is experiencing a financial bubble.

		All Data		Bubble	
Coef	Variable	Estimate	t-stat	Estimate	t-stat
γ_1	Total Net Asset	-0.638	-81.775	-0.265	-6.112
γ_2	Age	0.797	112.018	0.510	11.563
γ_3	Past return	0.342	13.776	-0.165	-2.551
θ_1	Herfindal Index	0.813	29.026	0.067	0.359
θ_2	Herfindal Index (norm'd)	-0.001	-4.196	0.042	8.847